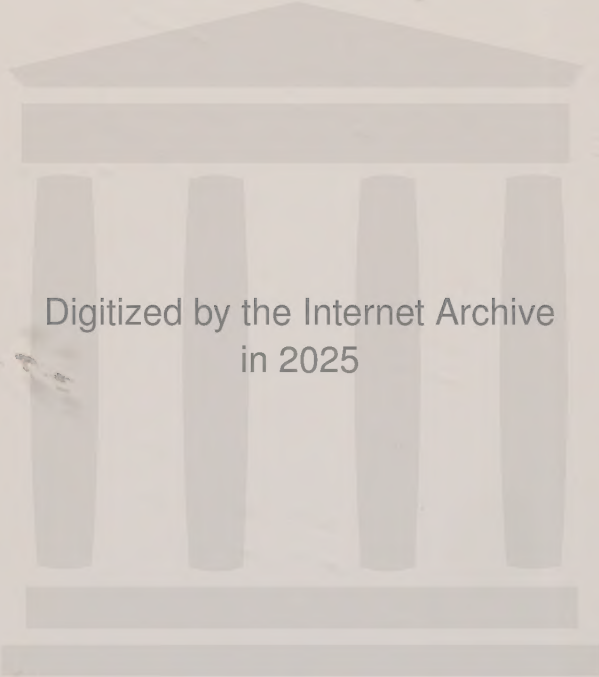




Digitized by the Internet Archive
in 2025

MEASUREMENT AND
ADJUSTMENT SERIES

EDITED BY LEWIS M. TERMAN

APTITUDE
TESTING

BY CLARK L. HULL

Professor of Psychology
University of Wisconsin



WORLD BOOK COMPANY

Yonkers-on-Hudson, New York
and Chicago, Illinois

WORLD BOOK COMPANY

THE HOUSE OF APPLIED KNOWLEDGE

Established 1905 by Caspar W. Hodgson

YONKERS-ON-HUDSON, NEW YORK

2126 PRAIRIE AVENUE, CHICAGO

Gideon chose his braves by their manner of drinking from a brook — a crude test of special aptitude, to be sure, but still a test. A French statesman of the Old Régime kept a gallery in which were hung paintings of battles and of domestic scenes. He would take young noblemen through his gallery and observe what subjects interested them most. Those who doted on battles he considered for court positions; and those who became wistful over domestic scenes he considered for military appointments. It is reported that his selections were in the main fortunate. Perhaps any test or no test would do in the presence of inspiration or genius. But for those who do and direct the world's work, day in and day out, only the most scientific of tests can be relied upon for the determination of aptitude. What has been done in the field of aptitude testing is shown, and what can be done is indicated, in this new book by an investigator who has made aptitude and its testing his special field of research

MAS: HAT-2

Copyright 1928 by World Book Company

Copyright in Great Britain

All rights reserved

PRINTED IN U.S.A.

TO BERTHA IUTZI HULL

SOURCE OF HEALTH, STRENGTH, AND TRANQUILLITY
THIS BOOK IS GRATEFULLY DEDICATED

25621

PREFACE

THE practice of psychological testing has advanced with great rapidity within recent years. The early crude methods are being replaced by scientific procedures, and the early naïve views in regard to the test-aptitude relation and the possibilities of tests are giving way before more adequate theories and more sober expectations. In a word, aptitude testing, like medicine and engineering, is ceasing to be a job for amateurs and is becoming the work of technically trained professionals.

It has been the purpose of the author to include within a convenient space two of the essentials of the training for aptitude work: (1) an account of the fundamental principles of aptitude testing and (2) an intelligible description of the most effective and the most economical methods of constructing batteries of aptitude tests. Specifically the book is designed as a text for university and college classes in aptitude testing and as a general handbook for those engaged in aptitude work of all kinds, whether in the form of vocational guidance, general personnel work, or employment selection.

The materials contained in the book have been gathered from many sources, through several years during which the writer has given instruction to university classes in aptitude testing. For much of the material on aptitude signs and much of the illustrative material on aptitude tests, he is indebted to students who have carried out investigations under his direction in the psychological laboratory at the University of Wisconsin. There has also been incorporated into the text the substance of a number of articles previously published by the author, on various aspects of aptitude testing. In practically all the chapters, however, will be found a considerable amount of original material not previously published.

Wherever matter from other writers has been utilized, suitable recognition is given in the place where each occurs. There are certain general obligations, however, which are so pervasive that a general statement is necessary. Of these, the influence of the writings of Truman Lee Kelley and of Godfrey H. Thomson is perhaps greatest. In a somewhat different sense there is a general obligation to the writings of Charles Spearman. In a more specific manner the author is indebted to Professor V. A. C. Henmon, and especially to Dr. Miriam West, who has read and criticized the more statistical chapters. The writer is greatly indebted to Professor A. S. Barr, who has read the entire manuscript and criticized rather minutely a large part of it. Mr. Selmer Larson has done considerable statistical work, including the construction of the table (Appendix I) for converting ranks directly into scores of amount on a linear scale. Mr. J. E. Caster made the drawings for a number of the pieces of apparatus and did the lettering for the numerous other figures. Bertha Iutzi Hull prepared the index and facilitated by constant encouragement the completion of the manuscript. Finally must be mentioned the invaluable assistance rendered by Professor Lewis M. Terman, who in his editorial capacity has given the entire manuscript a minute and painstaking criticism.

C. L. H.

MADISON, WISCONSIN

CONTENTS

	PAGE
EDITOR'S INTRODUCTION	xiii

PART ONE: PRINCIPLES

CHAPTER

I. INTRODUCTION	1
Psychological testing a judicious sampling of human behavior — Aptitude tests proposed in Plato's "Republic" — Aptitude testing and the rise of experimental psychology — Pioneer work of Cattell and Galton — The correlation coefficient brings order out of testing chaos — Alfred Binet and the Binet-Simon Tests — Otis and the army tests — Modern movement toward tests for specific aptitudes	
II. APTITUDE DIFFERENCES	21
People differ from each other — The law according to which individuals differ quantitatively from one another — Graphic representations of individual differences — The theoretical "normal" curve of distribution — How much better endowed is the best than the worst in an ordinary population? — How a person's traits differ among themselves — Some striking examples of asymmetrical mental development — Trait differences as revealed by educational achievement — Trait differences may be estimated on the basis of aptitude tests — Distributions of trait differences conform to the normal law — Trait differences about 80 per cent as great as individual differences	
III. VARIETIES OF TESTS	50
Educational proficiency tests — Trade tests — Trade tests <i>vs.</i> aptitude tests — Specific <i>vs.</i> general aptitude tests — Tests of general scholastic aptitude <i>vs.</i> tests of aptitude for general prudence in personal affairs — The number of test units employed in test batteries — Miniature tests <i>vs.</i> tests of abstract traits — Apparatus <i>vs.</i> non-apparatus tests — Individual <i>vs.</i> group tests — The conversion of individual tests into group tests — Some typical pencil-and-paper tests — Examples of group tests involving apparatus — Speed <i>vs.</i> power tests — Time-limit <i>vs.</i> work-limit tests — Tests of various stimulus-response mechanisms differentiated — Tests of sensory efficiency — Tests of motor efficiency — Tests of "mental" efficiency — Tests of character and temperament	
IV. ANATOMICAL AND OTHER ALLEGED SIGNS OF APTITUDE	111
Aptitude tests <i>vs.</i> aptitude signs — Typical claims regarding anatomical signs — Validity of character judgments based on photographs — Relation of academic aptitude to character	

judgments based on photographs — Have physiognomists special skill in judging character from photographs? — Validity of judgments of practical intelligence based on photographs — Evidence of blond and brunette coloring as signs of temperament — Validity of character judgments based on seeing the subjects in person — Evidence as to the significance of convex and concave profile — Dimensions of the head as signs of aptitude — The endocrine glands and somatico-behavior types — Kretschmer's somatico-behavior types — Naccarati's morphological index and academic aptitude — Does the shape of the hand reveal character? — Evidence as to the revelation of character by handwriting — Social aggressiveness and chemical constitution of urine — Chemical constituents of the blood and character traits — Heart rate and blood pressure as signs of aptitude — Heart rate, blood pressure, and emotional upset during testing

V. DERIVATION OF THE APTITUDE PROGNOSIS FROM RAW TEST RESULTS 156

Difficulties of choosing an aptitude-prediction scale — The critical test score — The coarse descriptive aptitude scale — Percentiles — Mental age and IQ — Scales based on range of group variability — Three methods of converting test scores into aptitude equivalents — Combining test scores by the individual psychograph — Combining test scores by simple summation — Combining tests on basis of estimated importance — The multiple-regression equation as a combining and forecasting formula

VI. THE BASIC CONSTITUTION OF APTITUDES AND TESTS . . . 184

Capacity *vs.* industry as gross determiners of success — Illustrative cases — Relative importance of capacity and industry as gross determiners of success — Are the able or the poorly endowed the more industrious? — The correlation of the interests of the individual with his abilities — Importance of fortuitous circumstances as gross determiners of success — The problem of generalization *vs.* specialization of traits — Spearman's theory of general and specific factors as aptitude determiners — Aptitudes as determined by specific factors — Aptitudes as determined by a combination of group and specific factors — Implications of the various theories of aptitude determination — A strict group-factor theory of aptitude determination — The theory of aptitude or behavior levels — Intellectual and motor abilities conceived as levels — Complexity of behavior as a factor

VII. FUNDAMENTAL RELATIONS AMONG APTITUDES AND TESTS . 218

Theory of group aptitude determiners — Theoretical determiner analysis of a concrete test-aptitude relation — How

common aptitude determiners produce correlations — Element of chance in determining size of correlation coefficients — The nature of attenuation — How more adequate measurement reduces attenuation — How true relations may be deduced from fallible data — The true validity of a test battery — Determiners of different sign, latent and negative correlation — Spurious correlation — Partial correlation illustrated — Limitations of partial correlation in determiner analysis

VIII. THE COMPOSITION AND YIELD OF TEST BATTERIES . . . 254

Two guiding principles in the final composition of test batteries — The increase in yield contributed by successive tests added to aptitude batteries — Principle of diminishing returns in the size of test batteries — Tests which can never make a perfect battery regardless of the number employed — Independent tests as fractional parts of perfectly predicting battery — Coefficient of multiple correlation not a perfect index of battery efficiency — “Per cent of subjects rightly placed” an ambiguous index of test-battery efficiency — The index of forecasting efficiency (*E*) — The increase in forecasting efficiency contributed by successive tests — Relation of the correlation coefficient to per cent of forecasting efficiency — Why the yield from test batteries is always limited

PART TWO: METHODS

IX. THE PSYCHOLOGICAL ANALYSIS OF OCCUPATIONAL BEHAVIOR 281

General outline of the six steps of test-battery construction — Step one: analysis of occupational behavior — The three parts of a completed psychological analysis — Analysis from observing the occupational activity — Analysis from performing the occupational activity — “Time study” methods as a means of analysis — “Motion study” technique adapted to occupational analysis — Experimental analysis in the psychological laboratory — Typical occupational analysis: “beamer” in textile industry — Typical occupational analysis: taxicab driver — Typical occupational analysis: free-hand drawing

X. THE ASSEMBLING OF A TRIAL BATTERY OF TESTS . . . 302

Intrinsic characteristics of desirable tests — Administrative characteristics of desirable tests — The utilization of tests already available — Some especially devised tests usually required — Pencil-and-paper tests: advantages — The completion form of test — The multiple-choice type of test — The true-false type of test — Relative effectiveness of recall, multiple-choice, and true-false tests — Influence of “guessing”

and "not guessing" — Sifting out the items composing pencil-and-paper tests — The arrangement of test items — Pencil-and-paper performance tests — Problems connected with the use of apparatus — Self-recording and self-scoring apparatus tests — Detailed procedure in assembling a typical test battery: free-hand drawing

XI. ADMINISTERING THE PRELIMINARY TEST BATTERY TO A TRIAL GROUP OF SUBJECTS 340

An adequate criterion score must be procurable — Number of subjects necessary for the trial group — The trial group should be similar to the group for which tests are intended — Test units must be in the form intended for the final battery — The order of tests in the battery — Experimental conditions — Securing coöperation from the trial subjects — The technique of timing — The duration of tests and batteries — The problem of scoring test results — Scoring pencil-and-paper tests by means of keys — Various types of scoring stencils — Machines for scoring multiple-choice tests automatically — Trial subjects, free-hand-drawing aptitude project — Experimental conditions, drawing-aptitude project — Typical test results and methods of scoring, drawing-aptitude project

XII. THE DETERMINATION OF THE ACTUAL APTITUDES OF THE TRIAL SUBJECTS 374

Three types of aptitude criteria — Five methods of scoring criteria — Scoring criteria by simple counting — Scoring criteria by objective measurement — Scoring criteria by ranking — Ranks not treated like ordinary units — Conversion of ranks into ordinary scale units — Concrete examples of scoring criteria by means of ranking — Scoring criteria by means of subjective scales — Characteristic defects of scores made by subjective scales — Method of rendering comparable, subjective scores by different judges — Rating scales — The man-to-man rating scale — Graphic rating scales — Ideal distribution of population on rating scales — Qualitative scales made from series of specimens — A criterion in detail: free-hand drawing

XIII. SELECTING THE FINAL APTITUDE BATTERY 421

Selection on the basis of correlation — Method of computing correlation coefficients illustrated — Systematic methods of computation necessary — The preliminary correlation work sheet — Use of the multiplication table — The standard deviation work sheet — The use of tables for finding powers and roots — The final correlation work sheet — How to guarantee absolute accuracy of the correlation coefficients — Correlation technique of the free-hand-drawing project — The problem of scoring errors in test performance — How to tell

CHAPTER

PAGE

whether scored errors will increase the strength of a test — How to score errors if retained for use in the final battery — General correlational characteristics desirable of tests for a battery — Significance of correlation signs in choice of tests — Four type correlational combinations considered as to battery desirability — Application of principles of final test choice to the free-hand-drawing project

XIV. COMBINING THE TESTS TO SECURE THE MAXIMUM FORECASTING EFFICIENCY457

If tests are not weighted, they weight themselves — A simple method of scientific weighting — Preliminary arrangement of data — How to write the normal equations — A system that insures accuracy in the computation of weights — An illustrative example of determining weights, four variables — Construction of the prediction formula, four variables — The formula in use — How to find the correlation yield (R) of a battery by formula — The correlation yield by direct computation — Limitations on the significance of R — Criterion estimates made by means of regression equation have narrowed dispersion — Practical difficulties arising from shrinkage in dispersion — A prediction formula which does not distort the criterion dispersion — A formula for forecasting on any desired scale other than the criterion — The free-hand-drawing project — The normal equations and their solution — The multiple-regression equation — The correlation yield (R), free-hand-drawing aptitude battery — The final check on the arithmetic — Practical defects of the multiple-regression equation as ordinarily used — Inconveniences of scientific weighting eliminated by a facilitating table — How to construct the facilitating table — A machine which makes aptitude forecasts automatically

APPENDIX

I. Table for converting ranks into linear scores	491
II. Table of $1 - r^2$ and $\sqrt{1 - r^2}$	493
III. Table of squares and square roots to 1000	495
IV. Multiplication table to 100×100	502

REFERENCES	523
----------------------	-----

INDEX	529
-----------------	-----

EDITOR'S INTRODUCTION

THE serious student of tests and measurement, whether his primary concern is with general intelligence, special aptitudes, or school achievement, is faced by the necessity of mastering a considerable body of fundamental principles and techniques. It is fortunate that for the most part these principles and techniques are the same in all the three fields mentioned, and the fact that they do have such wide application makes it all the more important that they be mastered.

Of course it is possible for one who has no scientific understanding of test methods to learn the simple procedures of test administration and test scoring, or even to carry on a certain amount of routine manipulation of test data, but there is a growing recognition of the dangers involved in this kind of work. The rapid multiplication of well-planned measurement courses in teachers' colleges and universities may be expected in time to remove all excuse for testing on the lower amateur levels. That time, in fact, seems to be drawing very near, and its advent is being materially hastened by the publication of textbooks which go beyond the superficial aspects of the subject.

Among the small number of texts which cover the ground of mental measurement in anything like a comprehensive manner, Professor Hull's *Aptitude Testing* is unquestionably one of the most significant. No other has dealt more searchingly with the fundamental issues involved in the construction and use of psychological tests, and no other excels it in directness and clearness of treatment. It is the work of an expert and embodies the results of many years of fruitful research. The student who has mastered this book will be effectually safeguarded from the innumerable pitfalls that beset the worker in this field, and will be in position to use test results with discrimination and insight. Although the book is con-

cerned most directly with the testing of special aptitudes, it is certain to find a wide field of usefulness as a basic text, inasmuch as the principles involved are for the most part common to the entire field of psychological and educational measurement.

Perhaps a word is in place regarding the purpose of Part II, which deals largely with the technique of test construction. The purpose is not that of preparing students to construct new batteries of tests, but instead to give them the information which they must have to enable them to select tests wisely for a given purpose and to interpret correctly the data which tests have yielded. With respect to this aim, the editor believes that *Aptitude Testing* brings together the minimum essentials for those who would make any really serious use of measurement methods in the classification of pupils for instruction or in their educational and vocational guidance.

LEWIS M. TERMAN

PART ONE

PRINCIPLES

APTITUDE TESTING

CHAPTER ONE

INTRODUCTION

THE most accurate method of determining the aptitude of an individual for a vocation or other activity is the test of life itself. The ultimate test must always be the learning of the vocation in the ordinary manner, after which we may observe the degree of the individual's proficiency when he has reached the limit of training. In some activities, such as tending certain automatic machines, the time required to discover a person's aptitude by the method of actual trial may be very brief. In other activities, such as those of the medical and legal professions, the period of training may be both prolonged and expensive. Indeed, in most important vocations the time and energy required to discover an individual's ultimate aptitude by the method of trial is so great that not more than one or two such determinations can be made in the course of an ordinary life. And if it turns out that the first choice of a vocation chances to be ill-advised, there is certain to be a great economic loss and an equally great loss in human happiness. A very similar situation exists in employment. The employee is likely to represent an economic loss to his employer for a longer or shorter period at the beginning of his service. In case the employee turns out to be ill adapted to the particular work, there results not only a net economic loss to the employer, but also a loss to the employee in happiness and self-respect.

Because of the magnitude of these various wastes and losses, there is arising a more and more insistent demand for expeditious and economical methods of discovering latent aptitudes. It is this necessity for economizing time and energy that has given rise to the modern short-cut

methods of aptitude prognosis by means of tests. This central fact must never be lost sight of. *A method of prognosis which is not at the same time reasonably quick and reasonably inexpensive has no excuse for existence.* These exacting conditions contribute much to the difficulties of the aptitude psychologist. Naturally, they determine to a considerable extent the technique which he is forced to employ.

Many methods of detecting latent aptitudes have been advocated. Some of these are known to be of considerable value, some have a little value, while it is to be feared that certain others have no value whatsoever. All are in process of experimental trial. Since, in all probability, the science of aptitude prognosis is in its infancy, it is reasonable to expect that many new and possibly revolutionary methods and devices may still be discovered. Certainly an enormous amount of experimental research is now being directed to this end. The wise worker in aptitudes will accordingly preserve an open mind to all possibilities. He will utilize any and every device or procedure that can be demonstrated scientifically to contribute in substantial amounts to the forecasting of aptitudes.

At present, various kinds of psychological tests are the chief means employed for making aptitude predictions. These tests range through tests of the higher mental processes, such as reasoning and ability to learn; tests of character and temperamental traits; tests of sensory acuity; tests of motor speed, coördination, and strength; tests of physiological traits, such as basal metabolism; chemical tests of the bodily secretions, such as of the blood and the saliva; etc. Though not tests, properly speaking, there may be included in the list of the aptitude psychologist's resources the determination of the various anatomical dimensions and proportions, past environmental conditions of the subject, and even the behavior and achievements of his ancestors.

PSYCHOLOGICAL TESTING A JUDICIOUS SAMPLING OF
HUMAN BEHAVIOR

Testing in all the applied sciences is performed on the basis of samples. The pure food authorities, for example, analyze a small quantity of milk chosen with some care from a can intended for babies, say, and assume with confidence that the remainder of the milk has the same chemical characteristics as those found in the sample. It would be a waste of time and energy to determine the actual amounts of the various constituents for the entire can of milk. Moreover, the milk would be consumed by such a procedure and none left for the babies. A very similar situation is found in as diverse a situation as the smelting of iron. There is no need to make an exact determination of the total amount of iron, manganese, and phosphorus, say, of an entire car of ore. Instead, a few ounces of ore are chosen at random from each of a number of specified places over the car. This sample is ground and thoroughly mixed. Of this only a small fraction of an ounce is finally analyzed. On the analysis of this minute sample the tons of ore remaining in the car are smelted. Again, if a bacteriologist wishes to discover the nature of a throat infection, he does not endeavor to examine all the bacteria the patient may harbor. He takes a sample from the patient's throat, from the sample makes a culture, and finally examines with his microscope a small sample from this culture. Or if a physician wishes to know whether a patient suspected of anemia is deficient in red blood cells, he does not attempt to count all the blood cells in the patient's body. He secures from the patient a few drops of blood and of this actually counts but a minute part. He assumes with confidence that the blood remaining in the patient's body has approximately the same proportion of blood cells as that examined under the microscope.

Human-aptitude testing is not essentially different from the application of tests in other sciences. The medium sampled differs, of course, from that in each of the other sciences, just as these differ from each other. The thing sampled in aptitude testing is in most cases human behavior. *A psychological test is the measurement of some phase of a carefully chosen sample of an individual's behavior.* We may, for example, determine how many times a person can tap with a telegraph key in one minute. This will be a sample of that subject's behavior under those particular conditions. Another sample taken an hour or a day later will not be quite the same as the first one. But if a suitable technique is followed, the successive repetitions of a test (after practice, fatigue, and such factors are allowed for) will not differ greatly from each other. Moreover, the average from several repetitions of a test will differ still less from other averages similarly obtained.

A rather different sampling of a person's behavior will be obtained from the disarranged-sentence test, as used by the United States army in the late war. The subject is presented with twenty-four disarranged sentences, such as:

oranges yellow are	true — false
noise cannon never make a	true — false
to life water is necessary	true — false
leaves the trees in lose their fall	true — false
forget trifling friends grievances never	true — false

He is directed to rearrange the sentences mentally so as to tell whether they are true or false, and then underline the word *true* or *false* at the right accordingly. He is to do this as rapidly and accurately as possible. The number of these sentences marked correctly in two minutes constitutes the test score.

The practice of taking samples of human behavior for purposes of examination is, of course, as old as formal educa-

tion itself. A school examination is merely a more or less well-chosen sample from a much larger number of possible reactions within a certain range, which the subject is presumed to be able to perform with some degree of success. Thus, a spelling examination or test may be a sample of fifty or one hundred words of varying difficulty. The ability of a given subject to spell this sample of words is, if the words have been well chosen, a distinctly useful index of the ability of the person to spell all the remaining words of the language.

APTITUDE TESTS PROPOSED IN PLATO'S "REPUBLIC"

The conception of specialized aptitudes and the desirability of having tests of behavior which will indicate in advance latent capacity, is very ancient. It appears repeatedly, for example, in Plato's *Republic*. Plato seems to have regarded it as of considerable importance in the conduct of an ideal state. In Book II we find Socrates leading the dialogue as follows:

"Really, it is not improbable; for I recollect, myself, after your answer, that, in the first place, no two persons are born exactly alike, but each differs from each in natural endowments, one being suited for one occupation and another for another. Do you not think so?"

"I do."

". . . From these considerations, it follows that all things will be produced in superior quantity and quality, and with greater ease, when each man works at a single occupation in accordance with his natural gifts. . . .

"But we cautioned the shoemaker, you know, against attempting to be an agriculturist, or a weaver or a builder besides, with a view to our shoemaking work being well done; and to every artisan we assigned in like manner one occupation; namely, that for which he was best fitted. . . . Now is it not of the greatest moment that the work of war should be done well? Will it not also require natural endowments suited to this particular occupation?"

"Then, apparently, it will belong to us to choose out, if we can,

that special order of natural endowments which qualifies its possessors for the guardianship of the state."

"Certainly it belongs to us."

"Then, I assure you, we have taken upon ourselves no trifling task."¹

Following this, Plato proposes as a means of accomplishing the task, that persons being considered for the military profession shall be given "actions to perform" which shall test the retentiveness of their memories, their power of resistance to deceptions, of resistance to timidity and fear in terrifying situations, and to the seductions of pleasure. Thus we find Plato sketching forth very definitely a set of tests for military aptitude. Some twenty-three hundred years later the dream conceived by the Greek genius was realized in the United States army mental tests.² Such is the halting course of progress.

APTITUDE TESTING AND THE RISE OF EXPERIMENTAL PSYCHOLOGY

The delay in any serious attempt to realize Plato's Utopian dream of having "each man work at a single occupation in accordance with his natural gifts" was inevitable. It had to await the development of an experimental and scientific psychology. This was slow to come. Possibly because of the paralyzing influence of metaphysical idealism in this field, psychology was one of the last of the disciplines to take on scientific form. The first psychological laboratory was founded by William Wundt at Leipzig in 1879.

¹ Translation by Davies and Vaughan.

² A striking fact indicative both of the genius of Plato and of the permanence of certain aptitude problems is that the proposal in *The Republic* to have a test of the ability of recruits to resist fear in terrifying situations was actually employed during the late war in a test battery for military aviators. Henmon (30), in his aviation tests, shot off a revolver near the subject and measured the amount of involuntary jerking of the hand and the disturbance of the heart rate which followed.

The early experimental psychologists had little interest in the problems related to the modern science of mental testing. They strove rather to discover general laws or principles of mental activity. They investigated the ways in which people are alike rather than the ways in which they differ. Nevertheless, methods of mental measurement and guiding principles of experimental procedure were gradually perfected. These were invaluable to the science of differential psychology which was to follow. It is upon differential psychology that aptitude testing mainly depends.

PIONEER WORK OF CATTELL AND GALTON

Among the men trained in Wundt's laboratory were a number of Americans. One of the first of these, J. McKeen Cattell, interested himself mainly in mental testing and the psychology of individual differences. As early as 1890 (*13*)¹ Cattell published an account of ten tests which he was already using. The list is characteristic of the time :

1. Strength of grip.
2. Rate of arm movement.
3. Two-point threshold on the back of the hand.
4. Amount of pressure required to produce pain on the forehead.
5. Least noticeable difference in weights.
6. Reaction-time to sound.
7. Time required to name ten colors.
8. Bisection of a 50-centimeter line.
9. Ability to reproduce a 10-second interval of time by tapping when subject thinks the interval has elapsed.
10. Auditory memory span for letters.

Some of these tests, particularly the more physical ones, had been in use in Francis Galton's anthropometric laboratory, established in England in 1884. Three of the tests of the above list are still in common use — strength of grip, color-

¹ Numbers in italic type, in parentheses, throughout the text refer to the numbered entries in the list of references on pages 523-528.

naming, and memory span. As a whole, they are characterized as being based on the simple rather than the more complex mental processes. This is a natural result of the fact that the early experimental psychologists confined their attention largely to the senses. Of the ten tests, only one, memory span, even approaches the complexity characteristic of many modern psychological tests.

To indicate the dawning state of aptitude testing at the time, we can do no better than quote Cattell's statement of the probable value of the tests of his list: "The results should be of considerable scientific value in discovering the constancy of mental processes, their interdependence, and their variation under different circumstances. Individuals, besides, would find their tests interesting and perhaps useful in regard to training, mode of life, or indication of disease." Clearly the interest is still chiefly in discovering more or less general principles or laws.

Francis Galton, in some remarks appended to Cattell's article just referred to, has the idea of aptitudes, as well as the technique ultimately to be followed, much more clearly in mind. He says: "One of the most important objects of measurement is hardly, if at all, alluded to here [i.e., in Cattell's article] and should be emphasized. It is to obtain a general knowledge of the capacities of a man by sinking shafts, as it were, at a few critical points. In order to ascertain the best points for the purpose, the sets of measures should be compared with an independent estimate of the man's powers. We thus may learn which of the measures are the most instructive." The context indicates, however, that Galton himself does not have in mind any very specific aptitudes, but rather some more or less generalized character traits or faculties.

During the next six years three other test series were put forward. All the proposed series were somewhat more psy-

chological than those of Cattell. They are also of special significance because of the eminence to which each of the three authors later attained. One was put forward (1891) by Hugo Münsterberg, a German working in the Harvard psychological laboratory. The second (1895) was by Emil Kraepelin in Germany. The third (1896) was by A. Binet and V. Henri in France. Later Münsterberg experimented extensively on special aptitudes and originated, among others, an aptitude test for street-car motormen. Kraepelin applied experimental methods to the study of mental pathology. Binet, assisted by Simon, devised for children the epoch-making tests which bear their names.

Meanwhile Cattell had begun (1894) to administer tests systematically to the students of Columbia College. This practice, after over thirty years, is still in operation at that institution in the form of Thorndike's tests for college freshmen. At first there was little or no attempt to connect the test results with the aptitude of the students for mastering the academic curriculum. By 1901, however, the aptitude aspect of the Columbia tests had become quite definite. In that year Clark Wissler published an account of the tests then in use, together with Pearson coefficients of correlation between many of the tests and scholarship as shown by university grades. The appearance of coefficients of correlation in aptitude work marks the beginning of a new era in testing.

THE CORRELATION COEFFICIENT BRINGS ORDER OUT OF TESTING CHAOS

For some time previously, scientists working with complex arrays of data had been in need of some uniform method of expressing the extent to which one variable, such as the height of fathers, agreed with a second variable, such as the height of sons. Francis Galton, who was especially concerned with this problem through his interest in evolution and inheritance,

invented a clumsy empirical method of determining the relations between two such variables. He would arrange the data on a square chart, stretch a silk thread across the center of the chart in such a way as to agree as well as possible with the centers of the various rows, say, and then measure the angle by which this thread deviated from the vertical. The tangent of this angle would then be looked up in a table. Since the tangent of a zero angle (showing no correlation) is zero, and since the tangent of the angle showing a perfect correlation (45°) is 1.00, this made a neat index or coefficient of the extent of the correlation.

Galton's idea was excellent, but his method was crude. No very accurate results could be hoped for which must depend upon the accuracy of a person's eye in placing a stretched thread. Accordingly, Karl Pearson, a British mathematician, devised a method by which Galton's tangent could be calculated directly and exactly without the use of the clumsy thread or even of the chart. Since the publication of Pearson's formula, Galton's function, when computed by means of it, has been known as the Pearson coefficient of correlation. This coefficient is always designated by the letter r . Because of the great importance of the coefficient of correlation, not only in the history of aptitude testing, but also for the adequate understanding of much of what is to follow in the present volume, a brief explanation of its significance will be given.

The correlation coefficient takes the form of a decimal which may range from $+1.00$ at one extreme to -1.00 at the other. A correlation of $+1.00$ means a perfect agreement or concomitant variation between two number series. An example of such an agreement is seen in the two series shown in Table 1. The larger the number in Column A, the larger is the number in Column B. It will be observed that it is not necessary for a perfect correlation that the

TABLE 1
SHOWING TWO SERIES OF NUMBERS PERFECTLY AND POSITIVELY
CORRELATED ($r = +1.00$)

A	B
1	4
2	6
3	8
4	10
5	12

numbers in the two series shall be of the same size, or that the range of the two series shall be the same. These matters as such have nothing whatever to do with the size of the coefficient. The series do correspond, however, in such a way that if any item in one be given, the corresponding item in the other can be determined with certainty. Thus, if a 6 should appear in Column A, it could be told with certainty that the corresponding value in Column B would be 14, 10 in A would correspond to 22 in B, and so on.

A correlation of -1.00 means a perfect *disagreement* in size between two number series. An example of such a perfect disagreement is shown in the two following series:

TABLE 2
SHOWING TWO SERIES NEGATIVELY CORRELATED ($r = -1.00$)

A	B
1	12
2	10
3	8
4	6
5	4

These are the same numbers as in Table 1. But in this second case the large numbers of A go with the *small* numbers of

B, and vice versa. It is particularly to be noted that in this case, exactly as in that of a perfect positive correlation, if any item in one series be given, the corresponding item in the other series can be determined with certainty. Thus, if a value of 7 appears in Column A, it will be certain to correspond to zero in B, and so on. This means that in aptitude work, a test giving a correlation, say, of $-.45$ with a criterion, has exactly the same prognostic value, other things equal, as a test correlating $+.45$. Perfect correlations, of course, are never obtained in aptitude work.

It follows that from the point of view of prognostic significance the opposite of a correlation of $+1.00$ is not a correlation of -1.00 but a correlation of $.00$ or zero. A zero correlation means that large numbers of one series tend to go neither with large nor with small numbers of the other series, as in the two cases shown above. Instead, the size of the numbers in one series will be related in a purely chance and *unpredictable* manner to the size of the numbers in the other series. In the following series (Table 3) the numbers are

TABLE 3

SHOWING TWO SERIES STRICTLY UNCORRELATED ($r = .00$)

A	B
1	10
2	8
3	6
4	4
5	12

the same as those used in the two previous examples. They are so placed, however, that the amount of agreement and of disagreement in the numbers of the two series is equally balanced, thus giving a zero correlation. In cases where the amount of agreement exceeds the amount of disagree-

ment, then there will result a positive correlation less than $+1.00$. In cases where the amount of disagreement exceeds the amount of agreement, then there will result a negative correlation varying between zero and -1.00 . Correlations between single test units and important aptitudes rarely run higher than $.45$ or $.50$. Test batteries rarely correlate with aptitudes higher than $.65$ or $.70$. For an explanation of the method of computing the correlation coefficient, see pages 423 ff. For further explanation of the significance of r , see pages 268 ff.

The invention of the method of computing the correlation coefficient came at a most opportune moment for the development of aptitude psychology. At the time of its publication, experimental studies involving tests were appearing at a rapidly increasing rate. Investigators were keenly interested in the relation of test scores to various criteria of aptitude and to each other. They were floundering helplessly in their efforts to judge the extent of correlation by the mere inspection of crude tabulations. One writer would claim correlation for data in which another writer could see no correlation whatever. Pearson's formula soon began to bring order out of this chaos. With this formula it was now possible for writers to state in definite and objective terms for any given set of data not only that correlations existed, but exactly how strong the tendency was. Wissler's study was one of the first important investigations to utilize the new tool. Without it, the science of aptitude testing in its present form would have been inconceivable.

ALFRED BINET AND THE BINET-SIMON TESTS

At this point we must note a special and very characteristic element of the aptitude-testing movement. We have already mentioned the fact that Alfred Binet proposed a series of psychological tests in 1896. He was connected with the

psychological laboratory of the Sorbonne. His interests at first appear to have been largely in general psychology, very much as were those of Cattell, but as time passed he also became more and more engrossed in the application of his tests to the determination of scholastic aptitudes. He differed from Cattell, however, in that he was concerned more with the general or average aptitude of children to master the curriculum of the elementary schools, and especially with the detection of those individuals conspicuously lacking this ability.

It is extremely interesting and by no means without significance that while Binet was working out his system of tests he was also patiently trying out all sorts of other indirect means of detecting latent capacity. Some of these methods would no doubt seem a bit bizarre to the academic psychologists of today. He measured the heads of children, apparently with an eye to the claims of the phrenologists. He made a painstaking investigation of the claims of graphologists (4) as to the revelations of character to be found in handwriting. He even had professional chiromnists examine the hands of selected children, exposed through a screen, to see if their mental traits could in that way be detected. In the end, all these investigations were abandoned as unpromising. The fact, however, that he was willing to try any method which held any promise whatever of serving his purpose is characteristic of the man's genius and originality.

In 1904 occurred an event which seems to have been a major factor in centering Binet's interests upon elementary scholastic aptitude. In that year he was appointed a member of a commission to investigate the mentality of children who were not making satisfactory progress in the elementary schools. A year later he published a tentative list of tests for children, and after three years (1908) his famous "scale

for measuring intelligence." Three years later, and shortly before his death (1911), Binet published a final revision of his tests. Few events have proved of as great significance in the development of psychology as the appearance of this scale.

Binet's tests consisted of a series of questions and commands which the person being examined was expected to answer or perform as the case might be. These test items were grouped according to their difficulty as determined by the proportion of children of different ages able to perform them successfully. In the 1908 scale there were groups of test items assigned to each year between III and XII inclusive. Those for year VI will give a general notion of the nature of the items:

YEAR VI

1. Knows right from left.
2. Can repeat sentence of sixteen syllables.
3. Can choose the prettier of each of three pairs of faces.
4. Can define familiar objects.
5. Can perform three commands all given at one time.
6. Knows own age.
7. Can tell whether it is morning or afternoon.

Almost at once translations and revisions of the Binet Tests began to appear, particularly in the United States. The first revision was made by Goddard. The most widely used of all the revisions up to the present time is probably that known as the Stanford Revision (86). It was worked out by Lewis M. Terman and a number of associates at Stanford University. Similar translations and revisions have been made in many languages. At present the tests are used more or less extensively in nearly every enlightened country of the world.

Perhaps the most unique contribution of Binet to testing practice is that of measuring mental development in terms of mental age. Indeed, the enormous vogue of the Binet

Tests is doubtless due in part to the simplicity and easy intelligibility of this concept, especially to persons little trained in technical psychology. If a child of eight years is able to go only as far up on the scale as the average child of six years, he will be said to have a mental age of six years. Since 6 is 75 per cent of 8, it is only a step to saying that he has 75 per cent of normal intelligence, or that he has an IQ of 75.

Curiously enough, this feature of the tests, which has contributed so much to their popularity, is in reality one of the weakest parts of the system. The weakness lies in the fact that the amount of mental growth from year to year, especially during adolescence, is by no means equal, but grows less and less until at last it actually becomes zero. This gross lack of constancy in the unit has caused no end of difficulty and confusion among investigators seeking to establish the Binet tests on a scientific foundation. One or two attempts have been made to eliminate the use of mental ages by substituting a point system. The Yerkes-Bridges Point Scale for Measuring Mental Ability (104) is an example. So far, no revision of the Binet tests which eliminates mental age as the measuring unit has received much support from the users of such tests. The persistence of the use of mental ages despite its defects as a measuring unit is doubtless due in part to the fact that in the earlier years of childhood the amount of growth is fairly constant. Fortunately it is this part of the scale that is used most extensively.

OTIS AND THE ARMY TESTS

The tests of the early aptitude psychologists were largely individual tests, as contrasted with group tests. The Binet tests follow this tradition. It was felt quite generally that group tests could not be sufficiently well controlled to yield results of much value. On the other hand, the enormous

economies possible by group methods, as contrasted with individual methods, led to occasional attempts to construct group tests. Historically, the most notable of these attempts was made by A. S. Otis. Shortly before the United States became involved in the World War, Otis had prepared a set of tests which not only could be administered to a large number of subjects at the same time, but which could be scored, by means of ingenious stencils, almost at a glance. This was most fortunate. When war broke out, the psychologists of the country found themselves confronted with the problem of testing enormous numbers of men with considerable accuracy, and yet at a speed previously unheard of. When they took stock of tests already available, they found many of their most difficult problems already solved in the Otis battery. With commendable generosity and patriotism, Otis at once made available to the government all of his materials. In the course of a few months the army psychologists under the direction of Robert M. Yerkes had evolved the famous group tests for military aptitude.

Two batteries of tests were evolved by the army psychologists — “Alpha” and “Beta.” Beta, the less important, was designed for use with illiterates and foreigners unfamiliar with English. It depended to a considerable extent on pictures. The directions were given mainly by pantomime. Alpha, the more important of the two, was designed for literates. It consisted of eight tests, as follows :

1. Directions
2. Arithmetic
3. Practical judgment or best answer
4. Synonym-antonym
5. Disarranged sentences
6. Number series completion
7. Analogies
8. General information

None of the test units of the Alpha battery required more than about 5 minutes to give, and some of them required only $1\frac{1}{2}$ minutes of actual working time. The total examination, including directions, required about 50 minutes. As many as 500 soldiers could be tested at one time by means of it.

During the war Army Alpha was given to over one and one-half million men. Such a huge volume of testing had never before been known. And, carried out as it was during intense public interest in everything pertaining to military affairs, the success of the undertaking gave psychological testing enormous publicity and prestige. As a result, the moment the war was over, there was a mad rush to transfer the testing methods found so useful in the army directly into industry, particularly in employment and personnel work. The attempt was made, often with little or no modification, to use the army tests to forecast aptitudes of the most diverse kinds. The endeavor to use tests designed for this one aptitude, to detect aptitudes of a very different sort, could have but one outcome. The result of this hasty and often ill-advised exploitation of an instrument really useful in its own field was temporary failure and disillusionment. There followed in industry a distinct reaction against aptitude testing.

The spectacular nature of the army-testing episode had at once a profound effect upon the aptitude-testing movement. Almost immediately a large number of pencil-and-paper group tests appeared on the market. For the most part, these tests were designed to detect and measure latent scholastic aptitude at one or another level. Nevertheless, nearly all of them bore a striking resemblance to the Otis and army tests in general form and structure. Under the pressure of war necessity, the technique of pencil-and-paper group testing had developed to a degree which ordinarily would have required many years.

MODERN MOVEMENT TOWARD TESTS FOR SPECIFIC APTITUDES

It has been and still is the custom to a very large extent, even among trained psychologists, to call nearly all psychological tests, tests of intelligence. Sometimes they are spoken of as tests of *general* intelligence. There is no exact agreement as to what this unknown something called "intelligence" is, though it seems to be thought of in the main as a kind of super-faculty. There are indications, however, that this idea of the function of psychological tests is falling into disrepute. Among testing experts there is a growing tendency to admit that what usually have been called "general intelligence" tests are in reality tests of scholastic aptitude; i.e., a kind of general average of the various aptitudes for learning the different school subjects. Unquestionably, these aptitudes constitute an extremely important group. Some of them resemble each other to a considerable degree. Some of them, however, evidently represent moderately distinct aptitudes. No very material advance in the testing of general aptitudes for mastering the common school curriculum is being made at the present time, either in the Binet type of test or the pencil-and-paper type of group test.

The recognition that if a test is to be of any particular value it must enable us to forecast a *particular* aptitude or group of aptitudes rather than measure some hypothetical or semi-metaphysical faculty, constitutes a great advance. During the period now happily drawing to a close, psychologists dominated by an essentially metaphysical notion of intelligence and consequently having no definite concrete criterion against which to test the validity of their tests, frequently moved in a circle in their scientific efforts. With the abandonment of this paralyzing idea of measuring *general* intelligence as the goal of testing activity, there is now appearing a vigorous and healthy concentration upon the develop-

ment of tests for the greatest variety of particular concrete aptitudes. Even in the relatively restricted field of scholastic aptitudes, numerous special-aptitude tests are being developed to forecast ability to learn particular subjects. These include handwriting, arithmetic, drawing, typewriting, shorthand, plane geometry, elementary algebra, Latin, modern foreign languages, and many others.

The list of non-academic aptitudes which have been attacked is already very extensive, and is rapidly increasing. The following partial list of workers who have been studied will indicate something of the scope of the investigations already begun :

General mechanic	Factory employee
Auto mechanic	Sewing-machine operator
Engine-lathe operator	Telegrapher
Motorman	Telephone operator
Auto driver	Compositor
Policeman	Grade teacher
Soldier	Registered nurse
Prison guard	Business executive
Fire fighter	Office boy
Aviator	Messenger boy
Retail salesman	General clerk
Traveling salesman	Comptometer operator
Insurance salesman	Hollerith operator
Department-store cashier	Free-hand draftsman
Restaurant waitress	Musician
Interpretative reader	

CHAPTER TWO

APTITUDE DIFFERENCES

APTITUDE differences fall into two distinct classes. From one point of view we consider how people differ from each other. The trait is constant; the individuals differ in the amount of it possessed by each. Such differences are called *individual differences*. From the second point of view we consider how the various traits or talents possessed by a single individual differ in amount. Here the individual is constant, but the magnitudes of his various traits differ. Such differences we may call *trait differences*. We shall consider these two types of aptitude differences in turn.

PEOPLE DIFFER FROM EACH OTHER

When considering individual differences in aptitude, we almost invariably find ourselves comparing extremes of ability—genius with subnormality. The mathematical powers of Newton or of Leibnitz may be contrasted with those of the moron scarcely able to make change. The inventive genius of Edison may be compared with the capacity of a man who is unable to adjust a simple lawn mower. Shakespeare's power of expression may be contrasted with that of the individual who, despite judicious instruction, can scarcely compose an intelligible letter. No matter to what realm of activity we turn, we are met by the same disparity of endowment, this same contrast between the very great and the very small in the way of aptitude. As a result of this universal tendency, scarcely any informed person at the present time maintains that people are equal in their aptitude for a particular activity or vocation.

Indeed, the most striking and significant things about people are their differences. It is due to these differences

that problems of vocational selection arise. If all people were of the same clay and cast in the same mold, there could exist no problem of choosing one person for a given service in preference to another.

THE LAW ACCORDING TO WHICH INDIVIDUALS DIFFER
QUANTITATIVELY FROM ONE ANOTHER

The subject of individual differences has been investigated very extensively for many years. As a result a great deal is known about this phase of differential psychology. One of the important results of this research is the discovery of the general law by which people differ, or vary, with respect to any given aptitude. This law is well established, not only for vocations but for the greatest variety of human behavior. It may be stated quite simply : In any given type of activity most persons tend to be of medium aptitude, or near it, on one side or the other. But passing upward or downward from this middle ability the number of persons at each step gradually grows less and less. At the extremes of great and small ability there comes at last a point beyond which no persons at all are found. This general law has been demonstrated experimentally to hold not only for behavior traits but for anatomical and many other traits as well.

An excellent illustration of the law in the anatomical field is seen in the distribution in height of an unselected population. Table 4 gives such a distribution. This shows very clearly that the great mass of English males fall either at 67 or 68 inches in height or close to it on one side or the other. But in passing upward or downward from this central stature the number of persons at each step grows less and less until at last no one is found who is taller than 77-78 inches or shorter than 57-58 inches.

TABLE 4

SHOWING THE FREQUENCY DISTRIBUTION OF ADULT MALES BORN IN ENGLAND (Adapted from Yule, *Introduction to the Theory of Statistics*)

HEIGHT IN INCHES	NUMBER OF MEN WITHIN EACH HEIGHT INTERVAL
56-57	0
57-58	1
58-59	3
59-60	12
60-61	39
61-62	70
62-63	128
63-64	320
64-65	524
65-66	740
66-67	881
67-68	918
68-69	886
69-70	753
70-71	473
71-72	254
72-73	117
73-74	48
74-75	16
75-76	9
76-77	1
77-78	1
78-79	0
Total 6194	

GRAPHIC REPRESENTATIONS OF INDIVIDUAL DIFFERENCES

It is often helpful in securing a unified view of a complex lot of data given by such a table to represent it by a simple curve or graph. Taking the data of Table 4 for purposes of illustration, the various heights of the men in inches are merely laid off on a base line, from the shortest to the tallest (Fig. 1). Then a small dot is placed above the stature intervals at a distance proportional to the number of persons

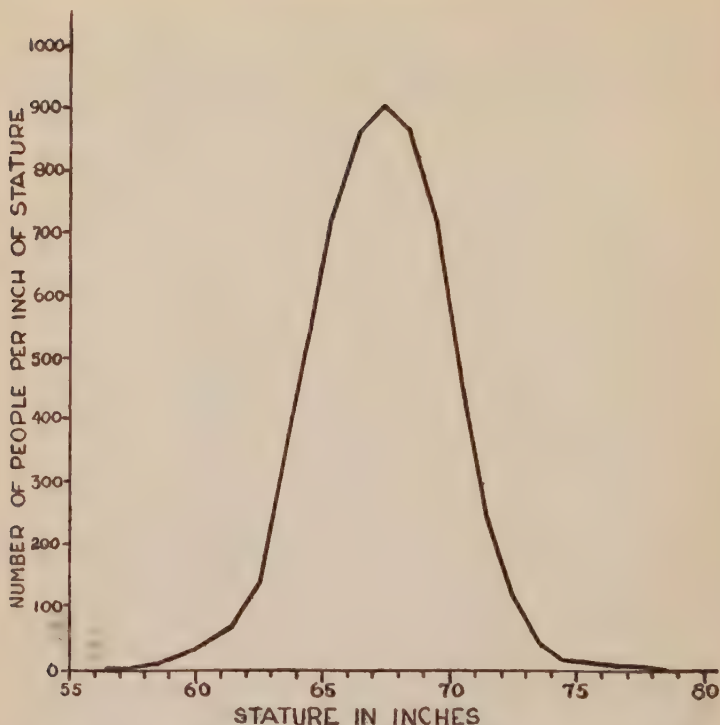


FIG. 1. Showing the distribution of stature among 6194 English males as shown in Table 4. (See Figure 2.)

found in each. The dots are then connected, usually by straight lines.

Sometimes a different method is used. The outlines of square columns are merely erected on the base line, each column to a height proportional to the number of individuals found in each interval (Fig. 2). The results of either method of representation show the distribution of statures not only to be symmetrical but distinctly bell-shaped, despite the use of straight lines.

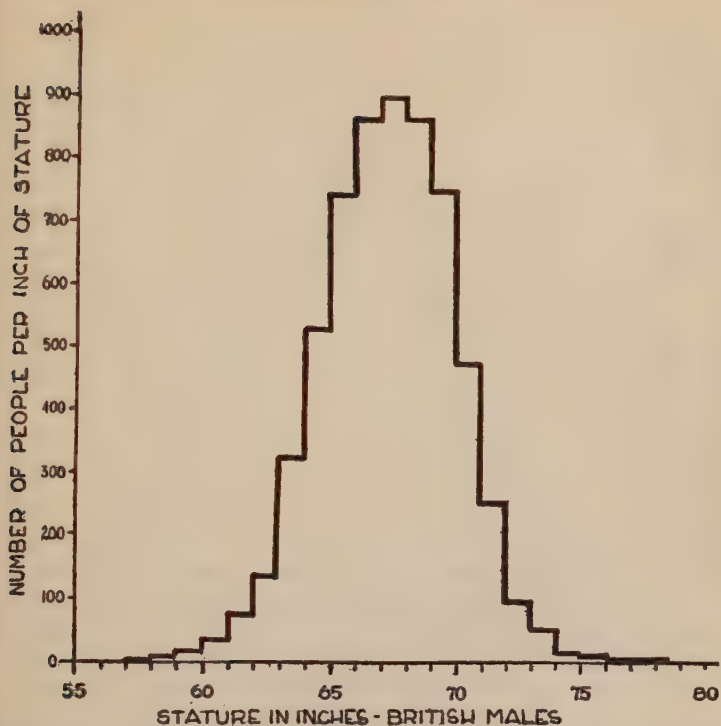


FIG. 2. Showing the stature distribution among the 6194 English males represented in Figure 1, but by means of the contour of a series of rectangles erected on the base line, each of a height proportional to the number of individuals falling within each interval of stature. (See Figure 1.)

Distributions revealing individual differences in respect to various forms of behavior are shown in Figures 3 to 7 inclusive. Owing to the relatively small number of individuals in some of these samples, the true form of the distribution is only imperfectly indicated, though the general outline in all cases can be recognized as approaching that of the theoretical bell-shaped curve. Hundreds of similar distributions

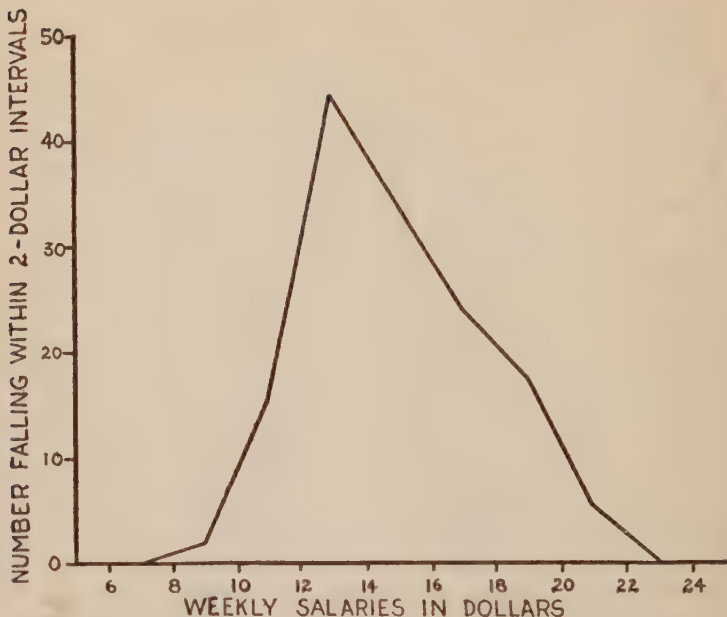


FIG. 3. Distribution of salaries of office boys. (Plotted from data published by Scott and Clothier.)

are scattered through the scientific literature, but those shown will suffice to indicate the essential nature of the distributions of most human abilities and aptitudes. Indeed, this bell-shaped distribution has been found so characteristic of all forms of human behavior that it is now regularly assumed as at least approximately true in any given case unless there is definite evidence to the contrary.

THE THEORETICAL "NORMAL" CURVE OF DISTRIBUTION

The particular shape to which all these distributions tend to approach is that known to mathematicians as the "curve of error." Such an ideal distribution is shown in Figure 8.

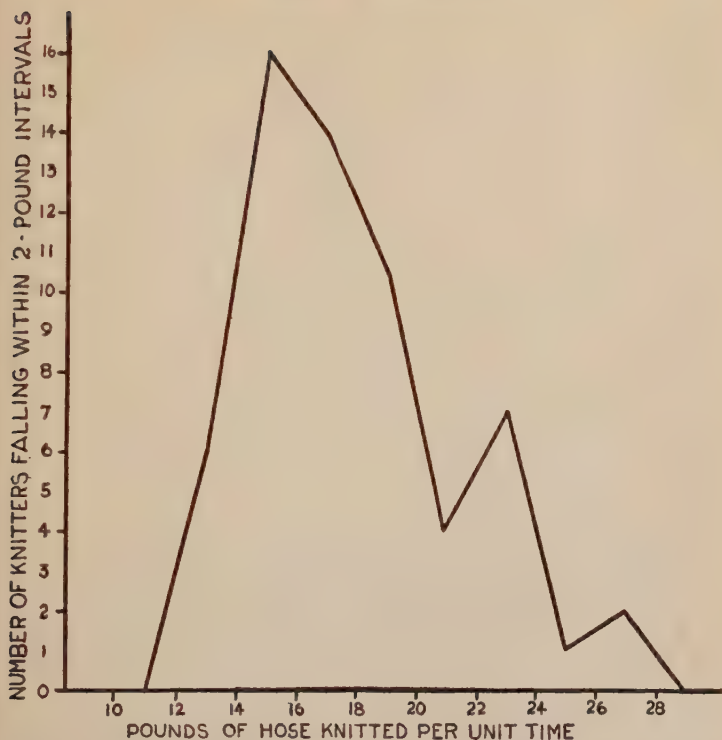


FIG. 4. Distribution of poundage output of knitting-machine operators. Plotted from data contained in an unpublished study of Howard Pollock (68).

This curve has been studied extensively by statisticians, and it has been found to possess numerous important characteristics. One of these concerns what is known as the “point of inflection.” A glance at the figure will show that the curve is relatively flat at the top. For some distance on either side of the middle it turns downward more and more. At a certain point, however, instead of becoming more and more steep, it reverses and begins to flatten out again. This

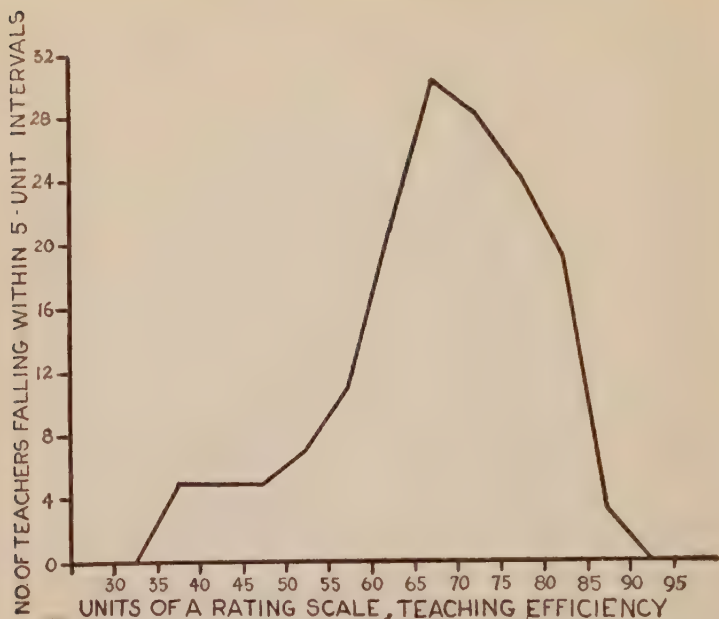


FIG. 5. Distribution of pedagogical efficiency of grade teachers as indicated by average rating by superiors. (Plotted from unpublished data received from Hertzberg and Thiel.)

point of change, or *inflection*, as it is called, is indicated (Fig. 8) by the letter *I*. Next to the position of the highest point of the curve (which falls at the mean) the point of inflection is the most readily located by mathematical methods. Probably for this reason the horizontal distance from this point of inflection to the line in the middle which marks the mean, has become the standard measure of variability or dispersion. Hence this distance is quite appropriately called the *standard deviation*. It is often expressed by the abbreviation "S. D.," or still more briefly by σ (the small Greek letter sigma).

In an ideally symmetrical distribution this distance is

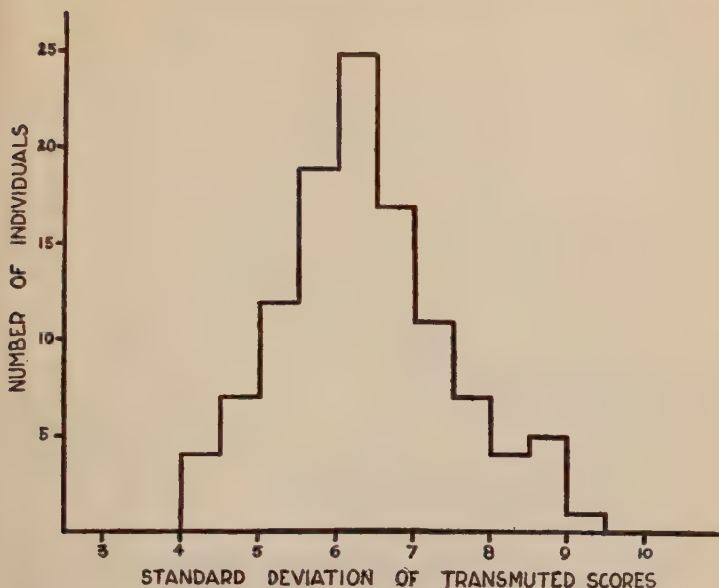


FIG. 6. Distribution of the individual variability on 35 tests by 107 first-year high-school students.

obviously the same, no matter from which side it is measured. That is to say, it will be 1σ or S. D. from the mean to the point of inflection on the right side and similarly 1σ to the point of inflection on the left side. Obviously the broader or the narrower any distribution may be, the longer or the shorter will be this distance. For this reason the standard deviation is an excellent index or measure of the amount of variability. While other measures of variability are sometimes used, none is so useful or reliable.

Mathematicians have discovered from computations based on the equation expressing the theoretical curve that approximately 68 per cent of the individuals in such a distribution fall between the two points of inflection. Taking the stand-

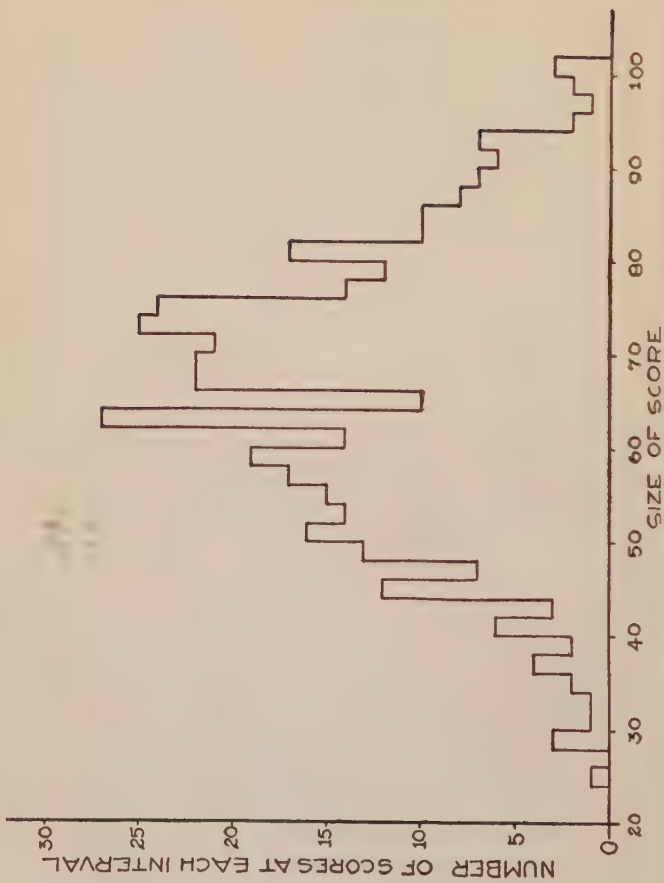


Fig. 7. Distribution of 400 trait scores constructed artificially by means of dice throws.
(See trait "0," pages 223 ff.)

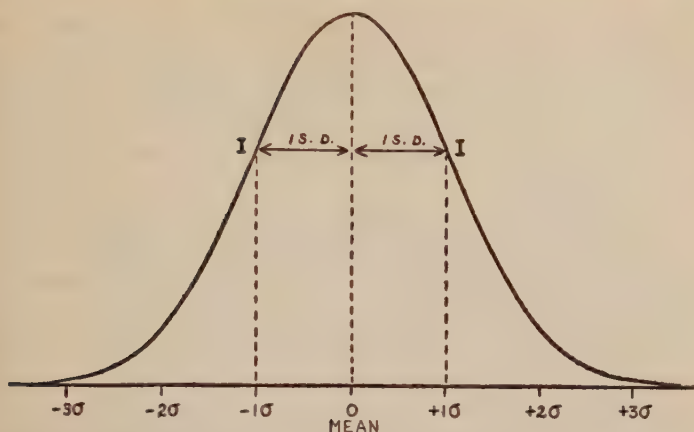


FIG. 8. An ideal distribution, showing the standard deviation (σ) in its theoretical relation to the mean and to the point of inflection (I).

ard deviation as a unit, similar values have been worked out for the various fractional distances from the mean. Arranged in the form of convenient tables, these are given in most standard works on statistics. Such tables show that practically all of the individuals of such a distribution are included within a distance of 3σ each side of the mean, or within a total range of 6σ in all. This may be seen very clearly by an inspection of Figure 8.

One important characteristic of distributions of human abilities as illustrated by Figures 1 to 6 is their evident continuity. There are no gaps. There is no break in the distribution of talent as we pass from one extreme to the other. The discovery of the strict continuity of the distributions of all known human abilities is of special significance because it serves to correct a rather widespread misconception. This erroneous belief is related to the so-called "psychological types." It is often naïvely assumed, for example, that any given person must be either intelligent or feeble-minded. In

thus sharply opposing intelligence to feeble-mindedness there is a tendency to assume that the two are qualitatively distinct, and by implication that there is an interruption in the continuity of distribution of human talent as the transition is made from one to the other. Nothing could be farther from the truth.

In a similar manner, people have been classified temperamentally as sanguine or melancholic and phlegmatic or choleric. It has been urged that there is a vital type, a motive type, and a mental type. Recently various types have been isolated as resulting from the supposed action of the endocrine or ductless glands (pages 138 ff.). Now there is no doubt that individuals may be found who will conform to any of the various alleged "types." It is clear, however, that there is no gap or discontinuity between the various alternative classifications represented by them. For the most part the "types" turn out to be nothing more than extremes of the distribution of individuals according to some special category. But in this sense the dwarf and the giant would also be types, since they represent the extremes in the distribution of stature. In all such cases the remainder of the population will be found scattered without a break from one extreme to the other, the great bulk crowding around a point about midway between.

HOW MUCH BETTER ENDOWED IS THE BEST THAN THE WORST
IN AN ORDINARY POPULATION?

Another important result which has come from the investigation of individual differences is an indication as to the magnitude of the differences which may exist in a given trait between the best and the poorest individuals in a normal population. Starch (82) reports the ratios existing between the best and the poorest performance on thirteen standard scholastic achievement tests by 36 eighth-grade children.

TABLE 5

SHOWING THE RATIO OF THE POOREST TO THE BEST SCORE OF 36 EIGHTH-GRADE CHILDREN ON THIRTEEN EDUCATIONAL ACHIEVEMENT TESTS
(Starch, *Educational Psychology*)

ACTIVITY TESTED	RATIO OF POOREST TO BEST SCORE
Reading (speed)	1 to 3.7
Reading (comprehension) . . .	1 to 3.5
Writing (speed)	1 to 1.6
Writing (quality)	1 to 1.5
Arithmetic (reasoning)	1 to 7.5
Arithmetic (addition)	1 to 15
Arithmetic (subtraction) . . .	1 to 8.5
Arithmetic (multiplication) . .	1 to 17
Arithmetic (division)	1 to 8
Spelling	1 to 2
Composition	1 to 2.3
Grammar	1 to 2.2
History	1 to 26

They are shown in Table 5. This table shows at one extreme that the best child on one study scored 1.5 times as high as the poorest. At the other extreme the most able child scored 26 times as high as the poorest. The middle value found by Starch shows the best of the 36 pupils to have scored about 3.5 times as high as the poorest.

Further evidence bearing on the same problem is furnished by the results of administering various psychological tests to groups of individuals. A typical series of this kind is afforded by some unpublished results secured by Charles E. Limp. He gave 34 standard psychological tests of the group type to 107 first-year high-school students. Owing to the presence of zero scores in a number of the tests, it is impossible to secure meaningful ratios for several of these tests. In the remainder, we find that at one extreme the best person scored 1.9 times as effectively as the worst, while at the

TABLE 6

SHOWING THE RELATION OF THE POOREST PERFORMANCE TO THE BEST PERFORMANCE OF 107 NINTH-YEAR PUPILS ON EACH OF 34 PSYCHOLOGICAL TESTS (Data furnished by Charles E. Limp)

NATURE OF TEST	AUTHOR OF TEST	LOW-EST SCORE	HIGH-EST SCORE	RATIO OF LOWEST TO HIGHEST
Information	Terman (group)	1	19	1 : 19
Best answer	" "	0	22	?
Logical selection	" "	0	16	?
Arithmetic	" "	0	18	?
Sentence meaning	" "	0	18	?
Analogies	" "	3	17	1 : 5.7
Mixed sentences	" "	0	16	?
Classification	" "	7	17	1 : 2.4
Number series	" "	0	22	?
Word meaning	" "	0	30	?
Motor reaction	Hoke (shorthand)	27	90	1 : 3.3
Speed of writing	" "	14	73	1 : 5.2
Quality of writing	" "	30	90	1 : 3
Speed of reading	" "	20	68	1 : 3.4
Immediate memory	" "	15	88	1 : 5.8
Spelling	" "	8	60	1 : 7.6
Symbols — speed	" "	14	75	1 : 5.6
Speed of decision	Downey (group)	2	10	1 : 5
Coördination of impulse	" "	0	10	?
Freedom from load	" "	2	20	1 : 10
Motor inhibition	" "	1	10	1 : 10
Volitional perseveration	" "	3	16	1 : 5.3
Interest in detail	" "	2	10	1 : 5
Motor impulsion	" "	0	3	?
Self-confidence	" "	1	10	1 : 10
Non-compliance	" "	0	9	?
Finality of judgment	" "	0	10	?
Easy directions	Woodworth & Wells	4	20	1 : 5
Cancellation (A-test)		37	82	1 : 2.2
Vocabulary		4	34	1 : 8.5
Dotting squares	Henmon	40	138	1 : 3.4
Arithmetic (addition)	Courtis	0	16	?
Arithmetic (multiplication)	"	0	14	?

opposite extreme the best person scored 19 times as high as the poorest. The middle ratio of Limp's results shows the best person as making a score 5.2 times as large as the poorest.

Somewhat more closely related to the practical problem of aptitudes are distributions showing the extremes of efficiency of persons who are actually engaged in gainful occupations. A number of these have been assembled from various sources. They appear in Table 7. The lowest ratio there shown is 1 to 1.4, whereas the highest is 1 to 5.1. The middle value falls at exactly 1 to 2.

TABLE 7

SHOWING THE RATIO OF THE LEAST EFFICIENT TO THE MOST EFFICIENT INDIVIDUAL ACTUALLY ENGAGED IN A VARIETY OF GAINFUL OCCUPATIONS

SOURCE	VOCATION	CRITERION	RATIO OF POOREST TO BEST WORKER
Loveday and Munro	Heel trimming (shoes)	No. pairs per day	1 : 1.4
Elton	Loom operation (silk)	Per cent of time loom kept in operation	1 : 1.5
Pollock	Hosiery maters	Hourly piecework earning	1 : 1.9
Wyatt	Loom operation (fancy cotton)	Earnings	1 : 2
Loveday and Munro	Bottom scoring (shoes)	No. pairs per day	1 : 2
Pollock	Knitting-machine oper- ators	Pounds of women's hose per hour	1 : 2.2
Scott and Clothier	Office boys	Weekly salary	1 : 2.3
Hertzberg and Thiel	Elementary teachers	Ratings of superiors	1 : 2.5
Farmer	Polishing spoons	Time per 36 spoons	1 : 5.1

The significance of the ratios given in Tables 5 and 6 is somewhat limited by the known fact that a person may make a very low or even a zero score on some of the tests in the list, yet possess an appreciable amount of the ability being measured. This tends to make the ratios in this table unduly large. On the other hand, the tendency to drop from employment all persons below a certain minimum of productivity prevents to a certain extent the appearance of extremely low efficiency scores among persons actually employed in gainful occupations. This tendency undoubtedly makes the ratios in Table 7 too small. The truth evidently lies somewhere between the 5.2 of Table 6 and the 2.0 of Table 7. We shall probably not be in great error if we conclude that *among individuals ordinarily regarded as normal, in the average vocation the most gifted will be between three and four times as capable as the poorest.*

An important practical corollary following from the above consideration is that the profits from conducting an industry may be very materially affected by choosing the workers from one or another part of the distribution of the aptitude in question. In many industries the same amount of floor space, tools, machines, etc., is required by a man from the lowest 5 per cent as by a man from the highest 5 per cent of the distribution of the particular ability involved. Yet such facts as those cited above indicate that men chosen from the highest 5 per cent would produce from two to three times as much per unit of overhead cost as men from the opposite extreme of ability.

HOW A PERSON'S TRAITS DIFFER AMONG THEMSELVES

We have just seen how extensively persons differ from each other in their aptitudes for success in any particular activity. Such differences are of special interest to the employer whose task it is to select from among numerous appli-

cants for a job the one best man to fill it. This is known as "vocational selection" and leads, among other things, to employment testing. We turn now to the consideration of the manner in which a person differs within himself in his aptitudes for success in the various possible activities or vocations. To distinguish these latter differences from individual differences, we shall call them *trait differences*. These latter facts interest primarily the person himself. Typically, this person is a youth who, at the beginning of life, seeks among the multitudinous vocations of the world that one in which he has the greatest aptitude for success. This second type of aptitude difference accordingly leads to the use of tests for purposes of vocational guidance. Despite the fact that trait differences have received scant attention in works on differential psychology, we shall try to show that they are for the masses of mankind of far greater significance than individual differences.

SOME STRIKING EXAMPLES OF ASYMMETRICAL MENTAL DEVELOPMENT

Just as in the case of persons differing from each other, there is no question in the minds of psychologists that each individual exhibits unevenness in the degree to which he possesses the various abilities. These asymmetries of endowment are frequently extensive. History contains many striking accounts of individuals who were talented in certain forms of activity, yet were comparatively feeble in others. The great mathematician La Place was appointed by Napoleon I to an important cabinet position, only to be found hopelessly incompetent. He soon had to be replaced. The great Napoleon himself was a wretched speller, and was so poor in Latin that his master considered him without intellectual gifts. Even his final examinations at the *École Militaire* were passed only "after a fashion." General Grant seems

to have been quite incompetent in business, and in music could not tell one tune from another. Dalton, the great chemist, was color blind. Oliver Goldsmith seems not to have been able to learn mathematics. The list could be extended almost indefinitely.

A special form of asymmetrical development is found in persons who are able to perform marvelous feats in mental computation. This ability is not infrequently associated with mediocre or even feeble intellect in other respects, though it is also met with in persons of excellent all-around ability. In the past a number of these mathematical prodigies, as they are called, have become quite famous.

The writer recently had the good fortune to observe rather intimately for several hours one of the greatest of these calculators now living. This man was somewhat sullen in disposition and rather uncouth in dress and general appearance. He was noticeably egotistical, especially in regard to his gift. He seems not to have finished the grammar school, but attributed his bad school record to his inability to get along with his teachers. His range of information on matters of ordinary interest was extremely limited. His mind seemed dull. His common sense was distinctly mediocre. As nearly as one may estimate the scholastic intelligence of an adult without a systematic test, his intelligence quotient appeared to be not above 80. Yet this man could give the product of any pair of three-place numbers instantly. Some notion of the magnitude of this feat alone, even as a memory performance, may be obtained if it is recalled that merely the condensed products of all the numbers up to 999 (as in Crelle's Tables) make up a volume about the size of a mail-order catalogue. He could give the products of four- and five-place numbers in two or three seconds. He factored and extracted roots of huge numbers with dazzling speed.

There is interesting corroborative evidence of this man's limited abilities in fields other than computation. In his stage appearances he was always accompanied by a clever stage partner who obviously managed the act. This stage partner also performed all the business transactions of the pair, such as bookings, transportation, baggage, etc. The writer was told by the late Harry Houdini, who knew the pair well, that the calculator received barely enough as his share of their joint earnings for his personal needs. It was added that sometimes the calculating genius would become sullen and would not perform. At such times the stage partner was in the habit of beating him into docility in a most shocking manner. At one time the mathematical prodigy rebelled and the partnership was discontinued for a few months. It is said that during this period he was quite unable to make a living even by the exploitation of his remarkable gift. In the end he was practically starved into returning to his iron-handed master.

As an example of a rather different type of asymmetrical development may be mentioned a self-styled "memory expert." For many years this man has been a conspicuous campus character of certain midwestern universities. He appears to have a rather remarkable facility in remembering the birth and death dates of historical personages. He makes his living by exhibiting his powers on street corners, challenging any one to give him a name for which he cannot give the dates. He rarely fails. Yet this man appears, to careful observation, to be largely ungifted in other respects. The writer was told by an intelligent woman who had been a classmate of his in the elementary school that this memory genius was quite the stupidest person in the class. He was the butt of the whole school.

As an example of a third and still different type of asymmetrical endowment, may be mentioned the case of the pickle

salesman recounted by Walter Dill Scott. It seems that President Scott had given some mental-alertness tests to a group of salesmen engaged in selling pickles to small retail grocers. When the test scores were compared with the sales records, he was astonished to find one of the very best salesmen with an intelligence rating almost of feeble-mindedness. At first an error was suspected, and so an investigation was made. It was found that the salesman in question was a large, fine-appearing man, with a hearty, jovial, friendly manner that made every one like him. He would go into a little grocery store, inquire of the grocer's wife with genuine interest whether Jimmy had recovered from the measles and how the baby was cutting teeth. Then he would talk pickles and they would buy. Yet this man was actually so lacking in ordinary clerical intelligence that it was necessary for the company to send a man with him on his trips to keep the records of his sales.

There is evidence that still other types of ability may exist in considerably more than average amounts in individuals who are otherwise definitely feeble-minded. Mrs. Hollingworth reports a case of a feeble-minded boy who possessed distinct talent in representative drawing. Into this category falls Blind Tom, who became famous a couple of generations ago for his ability to reproduce on the piano long musical selections after a single hearing. He seems to have been clearly feeble-minded in other respects. But such isolated observations, while interesting and suggestive, must be supported by more precise results obtained by scientific experimentation performed on a large scale. While this field has been by no means adequately explored by experiment, important results have been obtained.

TRAIT DIFFERENCES AS REVEALED BY EDUCATIONAL
ACHIEVEMENT

T. L. Kelley (43) gave the Stanford Achievement Tests to 96 Palo Alto eighth-grade children. He then made a very careful statistical examination of the data to discover the extent to which each individual child showed statistically reliable differences among his abilities on the nine school subjects included in the investigation. Kelley's final results are shown in Table 8. The figures in this table give the per cent of the 96 children showing significant difference between the scores of each child in each pair of school subjects tested. The first item, for example, indicates that 34 per cent of the children showed differences between their individual scores in arithmetic reasoning and in computation, in excess of what chance alone would produce. The large size of these values is truly impressive as indicating the extent to which asymmetries of development exist. Moreover, Kelley states that

TABLE 8

PERCENTAGE OF DIFFERENCES IN INDIVIDUAL TEST SCORES IN EXCESS
OF CHANCE, ASSUMING THE RELIABILITY OF EACH TEST IS .824
(Taken from Kelley)

	COMPUTATION	ARITHMETIC REASONING	WORD MEANING	SENTENCE MEANING	PARAGRAPH MEANING	LANGUAGE USAGE	SPELLING	SCIENCE INFORMATION
Arithmetic reasoning	34							
Word meaning	39	30						
Sentence meaning	40	26	13					
Paragraph meaning	35	26	14	17				
Language usage	39	33	24	23	17			
Spelling	31	36	28	33	25	30		
Science information	35	20	24	25	21	30	31	
History and literature information .	37	26	20	20	20	27	31	9

two thirds of the children showed one or more of these inequalities of development. These results are especially significant when it is considered from what a narrow range of possible behavior the above data have been taken. It is highly probable that if measures of musical, drawing, mechanical, athletic, and sales abilities had been included, the percentages of differences would have run much higher still.

TRAIT DIFFERENCES MAY BE ESTIMATED ON THE BASIS
OF APTITUDE TESTS

Granted that important differences exist among the aptitudes of individuals, the practical question arises as to whether these differences can be estimated in advance by means of specially constructed test batteries. Incidentally, if such differential prognosis can be accomplished, it will throw additional light on the nature and extent of such differences.

With the purpose of exploring the possibilities in this field, Hull and Limp (42) carried out an experiment with 73 first-year high-school students. Separate test batteries were made up with which to predict the aptitudes of individuals in (1) shorthand, (2) typewriting, (3) high-school English, and (4) high-school algebra. All of the students were tested by the four batteries, early in the year. At the end of the year the final school marks of each child in each of the four subjects of instruction were made up with great care.

As a preliminary control experiment these marks were correlated with the estimates of the corresponding aptitudes which had been made on the basis of the scores received by each individual on the four test batteries. The correlation coefficients obtained were .51, .61, .65, and .74, respectively. These coefficients mean that the four batteries were capable of forecasting their respective aptitudes at least as well as the average of special-aptitude test batteries.

To determine the extent to which the batteries were able to *differentiate* the four aptitudes in question *from each other*, the correlation was computed between (1) the *difference* between each pair of aptitudes, and (2) the *difference* between the corresponding estimated aptitudes made on the basis of the test batteries. The question was, if an individual's algebra-aptitude prediction is a certain amount higher than his shorthand-aptitude prediction, how strong is the tendency for the two actual aptitudes to differ in the same direction and to the same degree? The answer to this question is given by the correlation coefficients of Table 9.

TABLE 9

SHOWING THE EXTENT TO WHICH DIFFERENCES BETWEEN APTITUDE ESTIMATES MADE FROM TEST BATTERIES CORRELATE WITH DIFFERENCES IN THE CORRESPONDING ACTUAL APTITUDES (After Hull and Limp)

	DIFFERENCE BETWEEN ESTI- MATED APTITUDES IN SHORT- HAND AND TYPEWRITING	DIFFERENCE BETWEEN ESTI- MATED APTITUDES IN SHORT- HAND AND ENGLISH	DIFFERENCE BETWEEN ESTI- MATED APTITUDES IN SHORT- HAND AND ALGEBRA	DIFFERENCE BETWEEN ESTI- MATED APTITUDES IN TYPE- WRITING AND ENGLISH	DIFFERENCE BETWEEN ESTI- MATED APTITUDES IN TYPE- WRITING AND ALGEBRA	DIFFERENCE BETWEEN ESTI- MATED APTITUDES IN ENG- LISH AND ALGEBRA
Difference between aptitude in shorthand and typewriting .	+ .17					
Difference between aptitude in shorthand and English . .	—	+ .42				
Difference between aptitude in shorthand and algebra . . .	—	—	+ .42			
Difference between aptitude in typewriting and English . .	—	—	—	+ .50		
Difference between aptitude in typewriting and algebra . .	—	—	—	—	+ .31	
Difference between aptitude in English and algebra . . .	—	—	—	—	—	+ .23

It will be observed that these correlation coefficients, while positive, are distinctly smaller than those obtained by the direct prediction of the several aptitudes. In most cases, however, they are sufficiently large to indicate that genuine aptitude differences not only exist among activities apparently very similar, but that to a certain extent these aptitude differences may be prognosticated indirectly by means of specially designed test batteries. Had the aptitudes involved been as different as free-hand drawing, preaching, repairing watches, selling insurance, chemical research, and playing first violin in an orchestra, it can hardly be doubted that the differentiation of ability by means of the tests would have been much more effective.

A novel variation of the above experiment is reported by Limp (50). He used the same data as were used in the investigation just described; but instead of organizing two distinct test batteries, one to forecast typing aptitude and the other to forecast aptitude in English so as to see which aptitude prediction would be superior in any given case, he constructed a *single* battery that would predict for any particular person which aptitude would be superior and by how much. This was accomplished by means of a special prediction formula so constructed that if the forecast came out negative it would mean that the subject would probably excel in typing, whereas if it came out positive it indicated that the subject probably would excel in English. The numerical size of the prediction indicated the probable amount of superiority of the favored aptitude over the other, in units of the ordinary marking scale. Thus a forecast of $+5$ would indicate that the subject would probably receive at the end of the year a grade in English about five points better than his grade in typing. These "predicted" differences were found to correlate with the actual differences to the extent of $+ .50$, which suggests a distinct power of differentiating the aptitudes within the individual.

DISTRIBUTIONS OF TRAIT DIFFERENCES CONFORM TO
THE NORMAL LAW

Despite the obvious importance of trait differences as indicated by the evidence already cited, they have received very scanty attention from investigators as compared with that suggested by the voluminous literature on individual differences. A partial reason, however, is not far to seek. It is clearly a very much easier task to find 100 persons all fully trained on any single vocation than to find a single person fully trained on 100 vocations. Indeed, life is too short for a single individual to have reached the limit of training in more than one or two important vocations. This fact evidently precludes the possibility of attacking the problem directly through the measurement of occupational trait differences. There remains the indirect method of measuring non-vocational aptitudes.

One would like to have information concerning trait differences similar to that which we considered above in connection with individual differences. Specifically one would like to know (1) whether or not the distribution of trait magnitudes within the individual follows the normal bell-shaped curve of distribution, and (2) how extensive is the range between the average person's best and his worst aptitude. In an attempt to secure at least preliminary answers to the above questions, the writer carried out the following investigation (40). The subjects were 107 first-year high-school students. They were tested by means of 35 standard psychological tests involving a considerable variety of functions. In order to facilitate comparisons, the scores on the 35 tests were all converted into equivalent values. There resulted a rectangular table of 35 columns, one for each test; and of 107 rows, one for each subject. Now if the distribution of a *column* of data were to be plotted,

it would yield a graphic representation of individual differences with which we are already familiar. But if the distribution of a *row* of the data be plotted, we shall have a graphic representation of *trait* differences of the person whose scores occupy the row. This, of course, is what we are now seeking.

The trait distributions of various subjects were plotted from these data. The general contours of these graphs were somewhat irregular, owing to the small number of items (35) involved in each. Nevertheless, there appeared a distinct tendency to approach the characteristic shape of the normal probability curve. In order to secure a clearer picture of the distribution of trait magnitudes, the scores of certain subjects having approximately the same mean of talent and degree of talent variability were pooled. The resulting distribution is shown in Figure 9. The approximation to the shape of the normal curve is unmistakable. The same tendency was noted in several other distributions of a like nature. The indication is clear that the distribution of talent within an individual follows the normal law much as do the distributions of individual differences.

TRAIT DIFFERENCES ABOUT 80 PER CENT AS GREAT AS
INDIVIDUAL DIFFERENCES

The writer's investigation just described also furnished evidence as to the important problem of the extent or range of trait differences as compared with the magnitude of individual differences. Figure 9 has already shown the range of trait differences to be considerable. The standard deviations computed from the talent scores of each of the 107 subjects showed that some individuals were more than twice as variable as others. The distribution of these variabilities is shown in Figure 6. The average trait variability of the entire group shown in this distribution was found to be 6.33

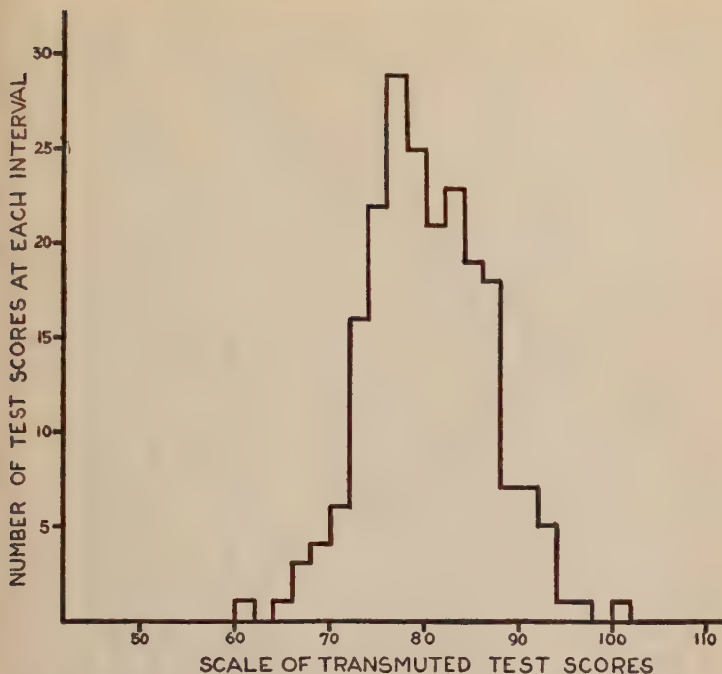


FIG. 9. Composite trait distribution of six subjects having approximately the same trait means and the same trait standard deviations, all being in the middle range.

points as compared with a variability among individuals of 7.00. In other words, the extent of trait differences within the average individual was found to approach rather closely the amount of difference found in a normal group in respect to any single trait. After a certain allowance had been made for inaccuracies of measurement, it was concluded that the trait differences here investigated average approximately 80 per cent as great as individual differences.

If subsequent investigation should show that the above results hold true of genuine vocational aptitudes, it will

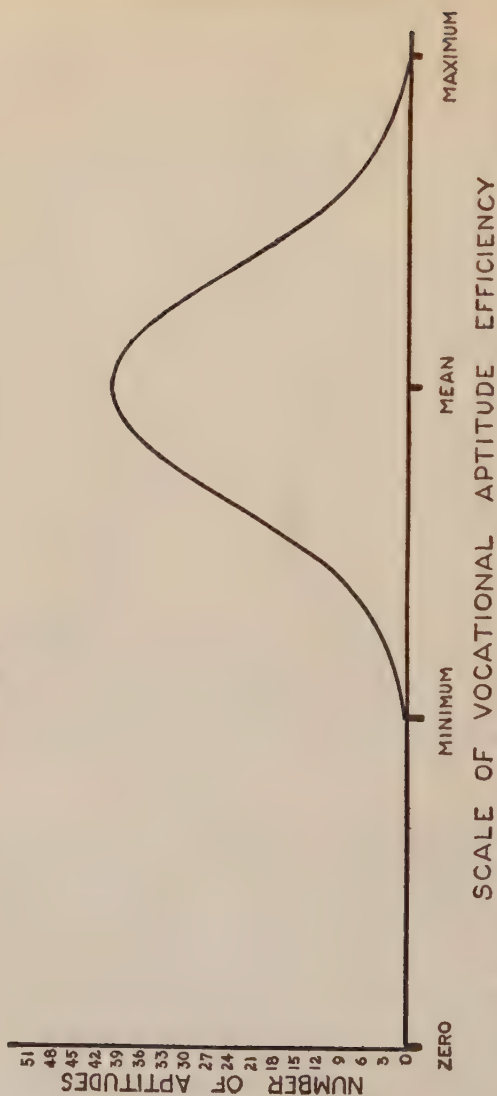


FIG. 10. Showing an approximate distribution of the average person's vocational-aptitude potentialities. In this figure the aptitude of maximum efficiency is placed at three times that of least efficiency.

mean that the average person has a great variety of occupational potentialities. Moreover, these potentialities are distributed according to the normal law, with a few very weak aptitudes and a few very strong ones, but with the great mass of a person's potentialities grouped about midway between. We have already seen reason to conclude that the best person in a normal group is between three and four times as efficient as the poorest. If the variability within the individual is 80 per cent as great as that, then the average individual's best vocational potentiality must be between two and one-half and three times as good as his worst. This general situation is represented graphically in Figure 10.

If anything closely approximating this should turn out to be true of vocational aptitudes, its significance as to the unrealized possibilities of vocational guidance would be profound. It means that if vocational choices are left largely to chance, as is evidently now the case, it will be very rare that an individual will choose the one vocation in which his aptitude is greatest. Indeed, by mere chance he would be about as likely to choose his worst. Such a mistake would be both a personal and a social tragedy of the first order. If we assume that in the chance choices of vocations the selections will be as likely to fall in the worse as in the better half of each person's vocational potentialities, the general average vocational efficiency must be not far from midway between the average person's best and his worst. It is the task of aptitude testing to forecast an individual's aptitudes in advance of choice, so as to be able to guide him into a vocation that will correspond as nearly as possible to his maximum potentiality.

CHAPTER THREE

VARIETIES OF TESTS

FROM the point of view of the purposes to which tests are put, they may be divided into two rather distinct classes: *aptitude tests* and *proficiency tests*. An aptitude test is a test designed to discover what potentiality a given person has for learning some particular vocation or acquiring some particular skill. A stenographic-aptitude test might be given to a boy who had never even seen a specimen of shorthand but who wished to know whether he possessed the capacity to become a successful court stenographer. A proficiency test, on the other hand, is a test designed to discover how perfect or skillful a person actually is in a given type of activity, quite regardless of the amount of training previously received. Such a test might properly be given by a prospective employer to an applicant for the position of stenographer, to discover whether she *now* has the desired speed and accuracy.

EDUCATIONAL PROFICIENCY TESTS

Proficiency tests tend to fall into two groups. In industry, proficiency tests are commonly known as *trade tests*. In the schools, on the other hand, proficiency tests are known as educational achievement tests or, more commonly, simply as *achievement tests*.

During the last decade there has been a great development in educational achievement tests. At present there are available proficiency tests for practically all the subjects taught in the elementary schools, for many of the subjects studied in the high schools, and for some even at the university level. Many of these, especially at the elementary-school level, are on the market. Some of the more important subjects, like reading, arithmetic, and spelling, are each

represented by a number of different achievement tests. In addition, there are scales for measuring proficiency (skill) in such diverse activities as English composition, handwriting, drawing, and sewing. It is said that attempts have even been made to construct a scale for measuring proficiency in baking cookies! A portion of an excellent handwriting scale is shown on page 411.

TRADE TESTS

We have seen (Chapter I) that the World War hastened enormously the development of group aptitude testing. In the same way it greatly accelerated the development of trade tests. The army needed at short notice large numbers of men skilled in a great variety of trades. The mere statements of the recruits as to their trade knowledge and skill were found to be of limited value because many men were inclined, either consciously or unconsciously, to exaggerate their proficiencies. The authorities accordingly assigned to a group of technical experts the task of constructing trade tests for the most of the important occupations of the army. The tests were designed to separate the men tested into four groups of proficiency: (1) novices, (2) apprentices, (3) journeymen, and (4) experts.

Trade tests tend to take three main forms, each of which has characteristic advantages and disadvantages. They are:

1. Verbal trade tests.
2. Picture trade tests.
3. Performance trade tests.

The verbal and picture trade tests are mainly tests of trade knowledge. As a rule, trade knowledge, as such, is not of very great importance. Such tests depend for their value mainly upon what tendency there may be for trade knowledge to indicate the presence of trade skill. The association be-

ORAL TRADE TEST (In part) ¹

MACHINIST AND MECHANIC. — Automatic Screw Machine Operator

COMMITTEE ON CLASSIFICATION OF PERSONNEL IN THE ARMY

Trade Test Division

Reproduced by permission of the Adjutant General

QUESTION 1

Q. When cutting steel where you want a high finish what kind of oil is used?

A. Lard-oil. Score 4

QUESTION 2

Q. On what kind of material do you use the highest spindle speed?

A. Brass. Score 4

QUESTION 3

Q. What are the two most common makes of automatic screw machines?

A. Brown & Sharpe (B & S).
Cleveland.
Gridley.
Acme.

Any two, Score 4

QUESTION 4

Q. How high should the parting tool be set in relation to the stock?

A. Center. Score 4

QUESTION 5

Q. What do you call the drill you use to start a hole with?

A. Spot (centering) (counter-sink). Score 4

¹ For the complete test see Chapman, J. C., *Trade Tests*, page 134. Reproduced by permission of Henry Holt & Co.

QUESTION 16

- Q. If a sharp forming tool is set below center what will it do?
A. Chatter. Score 4

QUESTION 17

- Q. When a die or thread tool cannot be used what do you use to form a thread in the rear of a shoulder when working brass?
A. Thread-roll (roller-die). Score 4

QUESTION 18

- Q. What kind of a drill do you use for brass instead of a twist drill?
A. Straight fluted (farmer) (flat) (gun).

QUESTION 19

- Q. In making a one-eighth forty fillister head screw, what size in decimals should the body of the screw be before threading?
A. .122 to .125. Score 4

QUESTION 20

- Q. What would you use instead of a hollow-mill to turn a part of considerable length?
A. Box *turner* (tool) (mill). Score 4

SCORING THE CANDIDATE

Score	Rating
19 and below	N
20 and 21	A -
22 to 50 inclusive	A
51 and 52	A +
53 and 54	J -
55 to 67 inclusive	J
68 and 69	J +
70 and above	E

There is no E - or E + rating.

ORAL TRADE TEST (In part) ¹

MACHINIST AND MECHANIC. — Automatic Screw Machine Operator

COMMITTEE ON CLASSIFICATION OF PERSONNEL IN THE ARMY

Trade Test Division

Reproduced by permission of the Adjutant General

QUESTION 1

Q. When cutting steel where you want a high finish what kind of oil is used?

A. Lard-oil.

Score 4

QUESTION 2

Q. On what kind of material do you use the highest spindle speed?

A. Brass.

Score 4

QUESTION 3

Q. What are the two most common makes of automatic screw machines?

A. Brown & Sharpe (B & S).

Cleveland.

Gridley.

Acme.

Any two, Score 4

QUESTION 4

Q. How high should the parting tool be set in relation to the stock?

A. Center.

Score 4

QUESTION 5

Q. What do you call the drill you use to start a hole with?

A. Spot (centering) (counter-sink).

Score 4

¹ For the complete test see Chapman, J. C., *Trade Tests*, page 134. Reproduced by permission of Henry Holt & Co.

QUESTION 16

Q. If a sharp forming tool is set below center what will it do?

A. Chatter.

Score 4

QUESTION 17

Q. When a die or thread tool cannot be used what do you use to form a thread in the rear of a shoulder when working brass?

A. Thread-roll (roller-die).

Score 4

QUESTION 18

Q. What kind of a drill do you use for brass instead of a twist drill?

A. Straight fluted (farmer) (flat) (gun).

QUESTION 19

Q. In making a one-eighth forty fillister head screw, what size in decimals should the body of the screw be before threading?

A. .122 to .125.

Score 4

QUESTION 20

Q. What would you use instead of a hollow-mill to turn a part of considerable length?

A. Box *turner* (tool) (mill).

Score 4

SCORING THE CANDIDATE

Score	Rating
19 and below	N
20 and 21	A -
22 to 50 inclusive	A
51 and 52	A +
53 and 54	J -
55 to 67 inclusive	J
68 and 69	J +
70 and above	E

There is no E - or E + rating.

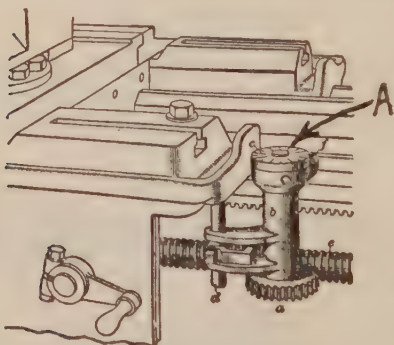
PICTURE TRADE TEST (In part)¹

MACHINIST. — Lathe Operator

COMMITTEE ON CLASSIFICATION OF PERSONNEL IN THE ARMY

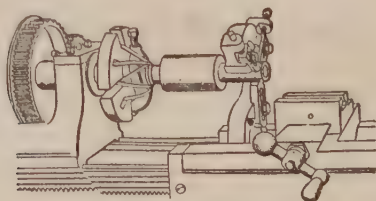
Trade Test Division

Reproduced by permission of the Adjutant General



PICTURE 11

14. Q. For what kind of work do you use the dial at A?
A. Cutting threads.

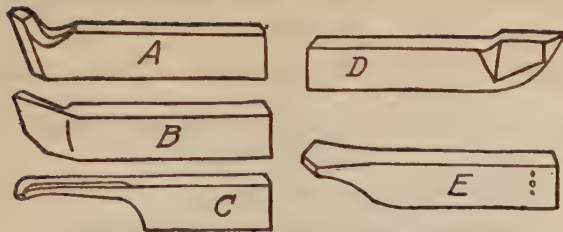


PICTURE 12

15. Q. How do you get the lace tight on the face plate?
A. Loosen face plate before lacing.

¹ For the complete test see Chapman, J. C., *Trade Tests*, page 225. (Reproduced by permission of Henry Holt & Co.)

PICTURE 13



16. Q. Name the tools in that picture.

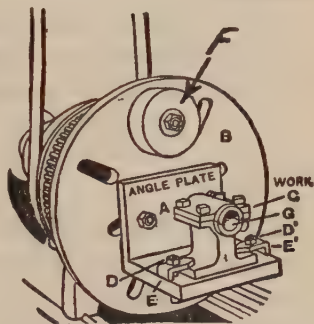
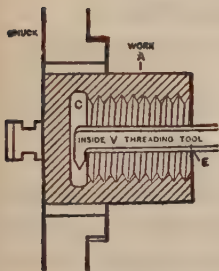
- A. (a) Diamond point.
 (b) Cutting-off (Parting)
 (c) Boring Tool.
 (d) Side Tool (Facing).
 (e) Centering Tool.

(All required)

PICTURE 14

17. Q. What is the purpose of the part F?

- A. Counterbalance
 (Balance).



PICTURE 15

18. Q. What is the use of the recess C?

- A. (1) Prevent breaking (jamming) tool.
 (2) Clearance.

(One sufficient)

RATING THE CANDIDATE

Score

Rating

0 to 16 inclusive.....	Novice
20 to 40 inclusive.....	Apprentice
44 to 60 inclusive.....	Journeyman
64 to 76 inclusive.....	Expert

SAMPLE OF PERFORMANCE TRADE TEST

(In part)¹

MACHINIST AND MECHANIC. — Lathe Operator

COMMITTEE ON CLASSIFICATION OF PERSONNEL IN THE ARMY

Trade Test Division

Reproduced by permission of the Adjutant General

TEST EQUIPMENT

Equipment:

- 1 Cast iron surface plate $7 \times 7\frac{1}{2}$ inches, weight approximately 10 lbs.
- 1 Cadillac steering spindle.²

Tools:

- 1 6 inch Outside spring caliper, Brown & Sharpe No. 806.
- 1 6 inch Inside spring caliper, Brown & Sharpe No. 807.
- 1 9 inch Combination square.
- 1 0 to 1 inch micrometer caliper with ratchet stop.
- 1 1 to 2 inch micrometer caliper with ratchet stop.
- 1 Pencil.

INSTRUCTIONS TO THE EXAMINER

- 1. Make certain that the TEST EQUIPMENT is complete and ready for the test.
- 2. Hand the candidate blue-print 6-L, No. 1.

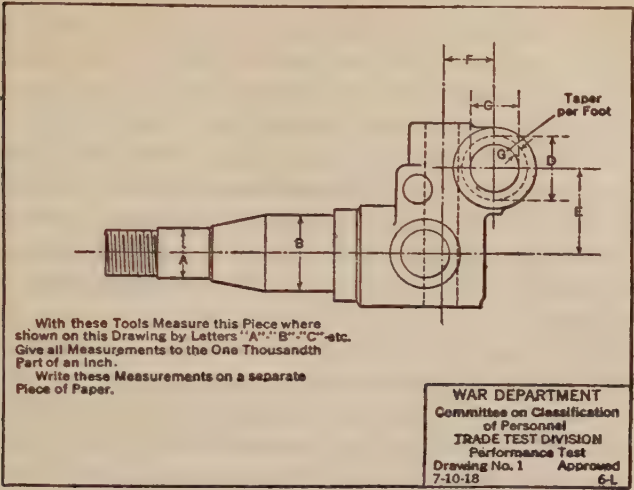
INSTRUCTIONS TO THE CANDIDATE

- 1. Say to the candidate: "Look at the instructions on this blue-print while I read them." Read distinctly and slowly all legends and measurements. Point to each thing as you read it.
- 2. Say to the candidate: "Are there any questions?"
- 3. Repeat, if necessary, all or any part of the instructions on the blue-print. Do not change them in any way.

¹ For the complete test see Chapman, J. C., *Trade Tests*, page 308. (Reproduced by permission of Henry Holt & Co.)

² Cadillac Price List of Parts, 1919. Part number KK 3063.

4. Answer any questions the candidate may ask during the test by repeating the instructions on the blue-print.



INSTRUCTIONS TO THE SCORER

- Score the candidate upon the Basis of the Measurements which he writes on the slip of paper supplied. All measurements are in inches. "Variations" in the **BASIS FOR SCORING** refers to variation from the correct measurement.
- Candidates will be scored in terms of variations from the correct measurements of the parts of the piece. A set of correct measurements will be furnished with each steering spindle.

INSTRUCTIONS TO THE RATER

- Rate a candidate's proficiency in his trade according to the following standards:

Score.....	Rating
25 or less.....	A
26 or more.....	E
- Rate a candidate J who receives
 - Rating A on this performance test and a
 - Rating J or higher on Picture Trade Test.

The Stenquist Mechanical Aptitude Tests (83), on the other hand, are apparently designed to test mechanical aptitude in general — i.e., a kind of average of a person's abilities to do all kinds of mechanical work. Here also fall the numerous so-called intelligence tests, such as the Binet-Simon Tests, the National Intelligence Tests, the Terman Group Test of Mental Ability, and many others. In reality these latter tests are essentially tests of general or average scholastic *aptitude* as contrasted with tests like the Orleans-Solomon Latin Prognosis Test, which is for a specific scholastic aptitude.

A third example of a test battery for general or average aptitude is Seashore's series of tests for musical talent. It would seem that these latter tests might be organized without much difficulty so that they would forecast independently the special aptitudes of singing, playing the violin, playing the piano, and many other special forms of musical talent. This has not yet been done. In their present state these tests merely give a measure of various kinds of behavior supposed to be involved more or less in all kinds of musical performance. As a result, the tests are taken by most users to indicate a kind of average potentiality for musical achievement.

TESTS OF GENERAL SCHOLASTIC APTITUDE *vs.* TESTS OF
APTITUDE FOR GENERAL PRUDENCE IN PERSONAL AFFAIRS

The best developed of the general-aptitude tests are those of average scholastic aptitude. An enormous amount of confusion has existed as to the essential nature of these tests. This may be due in part to the rather general custom of calling them "intelligence" tests without qualification. Our inveterate tendency to think in terms of the faculty psychology, even though we have long repudiated it verbally, has doubtless contributed much to the confusion. Moreover,

it must be remembered that the psychologists who have evolved the so-called intelligence tests have almost without exception been men intimately connected with schools. Quite naturally they have considered the aptitude to master the academic subjects taught by the schools as the one essential kind of intelligence or talent. It is not accidental that the validity of all these tests, beginning with the original Binet Scale, has been judged ultimately by the criterion of academic success — school marks and the ratings of teachers as to intelligence.

It is true, of course, that if a child is very strikingly unable to perform the tasks required by the lower schools he will not be very likely to have much success later in the pursuits of everyday life. This general fact is well established, though there are notable exceptions. No doubt this correlation between academic success and success in life is due to the fact that the abilities required for success in the common school subjects are in part the same as those required for success in many of the varied activities of daily existence.¹ It has consequently come about that tests primarily designed and employed to test average scholastic aptitude have found a certain usefulness in identifying individuals who, as adults, will be unable to manage their affairs with sufficient prudence to make a satisfactory living.

This kind of aptitude — not very different from what we call common sense — is probably one of the most composite of all aptitudes. It would accordingly seem desirable, for purposes of forecasting aptitude for successful management of personal affairs, that a much wider range of behavior should be sampled than is done by the present scholastic-aptitude tests. It is probable that if there were added to the largely verbal elements of tests of the Binet type some

¹ For a systematic discussion of the nature of aptitudes and intelligence, see Chapter VI.

tests of manual and especially mechanical nature, the combination would have considerably increased value in predicting the aptitude for prudent individual management or economy.

As an indication of the general confusion on this point, the following incident may be mentioned. Each of a number of prominent psychologists was asked by a psychological journal to write out a statement of what he considered general intelligence to be. All were sure that it is something very important, but no two agreed as to exactly what.

It will be explained in a later chapter (pages 201 ff.) that it is doubtful whether any unitary faculty of general intelligence exists. We have, instead, a large number of more or less specialized potential aptitudes or *intelligences*. When speaking of intelligence we should, therefore, if we would be quite accurate, speak in the plural rather than the singular. If the above contentions are true, it is evident that a "general" intelligence test could be nothing but a kind of general average of all the potential aptitudes for any given person.

It is probable that the future will see some approximation to a really general or universal intelligence test by means of which the various type aptitudes of an individual may be *separately* forecast. The test would be general in the sense that a single test would be administered, but the numerous aptitude forecasts derived by uniting the various test scores in different combinations would be highly specific. If desired, of course, an average of all these distinct aptitude estimates could be taken. This would probably be as close to a truly "general" intelligence test as will ever be attained. *General intelligence is thus defined as a kind of average of all possible aptitudes.* The natural objection to such an average aptitude rating would be that, being so general, it would yield an exceedingly blurred and indistinct indication of anything in particular. This is true. After all, it is usually

the particular in which we are interested. If each of the important type aptitudes were known, it is doubtful whether we should have much use for the general average of all.

It must not be inferred from the foregoing that special aptitudes are simple in the sense that they involve the ability to perform but one kind of act. Probably no strictly specific aptitude exists. We are dealing here merely with varying degrees of generality in the sense of compositeness. The vocational activities which correspond even to the most specialized of aptitudes are to be thought of as a kind of joint function or average of a person's ability to perform a *limited* number of fairly distinct activities. We may consider, as an example, the duties of a telegraph operator. Even a superficial analysis of this work reveals the fairly distinct duties of receiving and sending. While doubtless not entirely distinct abilities, they are probably by no means identical. The aptitude of a telegraph operator is thus a kind of average of at least these two abilities. The most of the aptitudes which would be called specific or special are probably more composite still, being specific only by comparison.

THE NUMBER OF TEST UNITS EMPLOYED IN TEST BATTERIES

A second method of considering aptitude tests relates to the number of test units employed for a given prognosis. The contrast lies primarily between the use of a single test unit and a battery of test units. It will be pointed out in a later chapter (pages 256 ff.) that most aptitudes are of such complexity that a single test unit will rarely be able to sample enough of its determining factors to make a useful prediction. This conclusion, based on theoretical considerations, is very generally confirmed by testing practice. There is hardly an aptitude test composed of a single test unit in use at the present time. Batteries are practically the universal form of aptitude test. The number of test units employed de-

pend in part upon the amount of time considered available for giving the test, but mainly upon the number of really useful test units available.

An example of a single test unit which is sometimes used for aptitude purposes is Seashore's phonograph test of pitch discrimination. It is significant, however, that this test is really but a part of a battery consisting of several units. Since it is the most important single unit of the battery, it naturally is used in preference to any of the others in cases where it is not convenient to use the entire series. Indeed, wherever a single test unit is employed for aptitude purposes, the phenomenon may be regarded as merely one of transition. Even with the most potent single units it is probable that additional tests, if properly chosen, will tap certain additional determiners of success in the vocation not sampled by the dominant test unit.

There are two special forms of test which may appear superficially to be exceptions to the general rule that single test units have no permanent place in aptitude testing. One of these is what has been called the "omnibus" test. This is a form of pencil-and-paper test made up of a large number of test items of several distinct kinds, all mixed together. The mixture of the various types of test items is not a mere jumble or haphazard arrangement, however. Instead, the various kinds of test items are arranged systematically in recurring cycles. For example, if it were decided to form an omnibus test from a combination of tests involving (1) common sense, (2) opposites, (3) analogies, (4) general information, and (5) synonyms, the first five items of the test would be made up of one item of each of the above types of test in the order listed, the next five would be a second series or cycle arranged in the same order as the first, and so on. Ordinarily there will be a dozen or more cycles.

By way of concrete illustration there is represented on page 65 two cycles of an omnibus test made up from the five test

EXAMPLE OF A FIVE-PROCESS OMNIBUS TEST IN THE MULTIPLE-CHOICE MODE OF PRESENTATION

DIRECTIONS. In each of the following statements you have four choices for the best ending. Before each form of ending there is a number. Read each one through and then put in the square at the right the number of the ending that makes the best sense. If you are not sure, guess. The first one is marked as it should be.

1. Cats are useful animals because (1) they are gentle
(2) they scratch (3) they catch mice (4) they are
afraid of dogs..... ☒ 3
2. The opposite of *high* is (1) lofty (2) low (3) good
(4) deep..... ☐
3. *Gun* is to *shooting* as *knife* is to (1) cutting (2) run-
ning (3) bird (4) iron..... ☐
4. America was discovered by (1) Drake (2) Hudson
(3) Balboa (4) Columbus..... ☐
5. A word meaning the same as *true* is (1) good (2) false
(3) accurate (4) beautiful..... ☐
6. Streets are sprinkled in summer (1) to make the air cooler
(2) to keep the dust down (3) to kill germs (4) to
give some one a job..... ☐
7. The opposite of happy is (1) sad (2) gay (3) ugly
(4) poor..... ☐
8. Ear is to hearing as eye is to (1) table (2) hand
(3) seeing (4) playing..... ☐
9. The most prominent industry of Detroit is (1) automo-
biles (2) brewing (3) flour (4) meat packing.. ☐
10. A word having the same meaning as *hope* is (1) conscience
(2) despair (3) aspiration (4) heaven..... ☐
11. Etc.

forms listed above, all having been adapted to the multiple-choice mode of presentation. The method of construction characteristic of the omnibus form will at once become clear if the reader will take the trouble to identify the several types of tests as they succeed each other.

One of the limitations of the omnibus type of test is that the instructions must be given at the beginning of the test period, once and for all. Where a number of highly distinct types of test activity are to be included in the cycle, and especially where the instructions for each form happen to be fairly complicated, the subjects are likely to get the instructions for the different tasks badly mixed if they are given all at the same time. A fairly successful method of meeting this difficulty is to give a practice exercise preceding the test. During this fore-exercise the instructions for each kind of test activity found in the cycle are given separately and are followed immediately by a brief practice before another type of test activity is taken up.

One of the chief advantages of the omnibus form of test is the extreme simplicity of its administration. For example, in place of the confusion and waste of time from repeatedly giving starting and stopping signals in administering a battery of five tests, the omnibus technique requires but one starting and one stopping signal and but a single time interval to be noted. It is also to be observed that it matters little where the stopping signal interrupts the work of the subject. Owing to the cyclical arrangement of items, a subject will have spent practically the same proportionate amount of time on each form of test activity as if he had been interrupted at any other time or place in the series. This fact would make it possible to give differential weighting to the various forms of test activity included in the cycle.¹ Up to the present this seems never to have been attempted.

¹ An explanation of the weighting of tests is found on pages 179 ff. and pages 459 ff.

A second variety of test unit which tends to be self-sufficient for aptitude purposes is the interesting form known as the "miniature" test. As contrasted with the systematic conglomerateness of the omnibus test, the miniature type of test is a true organic test unit. It will be convenient to discuss this important test form in connection with its antithesis, the class of tests designed to measure abstract traits.

MINIATURE TESTS *vs.* TESTS OF ABSTRACT TRAITS

Aptitude tests may be divided into (1) those which attempt to duplicate all in one test the essential activities of the occupation, and (2) those which are designed to isolate and measure separately the component traits supposed to constitute the determiners of success in the aptitude in question. The first type is appropriately called "miniature" test because it really is a kind of miniature of the aptitude activities. The apparatus is usually reduced in size for convenience of housing and is so constructed that various aspects of the subject's behavior may be recorded in a convenient manner. The second and contrasting type of tests may appropriately be called "tests of abstract traits."

As an example of a miniature test may be cited the Wisconsin miniature test for engine-lathe aptitude. This apparatus is shown in Figure 11. It attempts to duplicate that part of an engine lathe which controls the movement of the cutting tool. To this end the arm *A* supporting the point *P* is mounted in such a way that the latter may be moved anywhere about the vulcanite plate *X* by the joint action of two screws placed at right angles to each other and turned by the cranks *H* and *H'*. It will be observed by Figure 11 that these cranks are duplicates of those actually used on an engine lathe. The task of the subject is to move the point *P* around the series of six electric contacts shown in plate *X* as rapidly as possible while going as directly as possi-

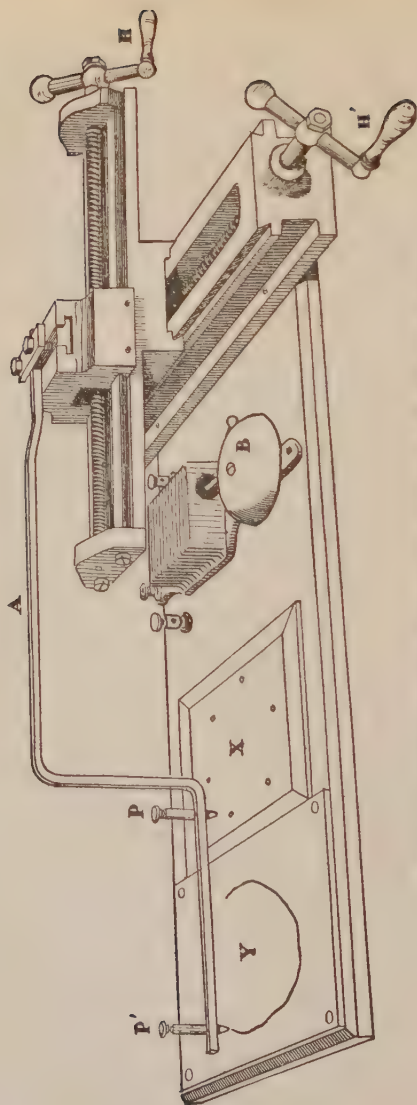


FIG. 11. Showing the Wisconsin miniature test for the engine lathe.

ble from one to the next. When the point touches a contact, an electric bell *B* rings to notify the subject that he has succeeded in touching it and that he may proceed to the next contact. In order to secure a graphic record of all of the movements of the subject in passing around the contacts, arm *A* is extended over area *Y*, upon which is placed a piece of paper. On this extension is a point *P'* which contains a small lead pencil which traces a duplicate of the path made by point *P*. A typical record is shown in Figure 11.

Some of the most ingenious examples of miniature tests have been devised to detect latent aptitude for motormen on the cars of electric railways. A glimpse of one of the most suggestive of these in respect to duplicating the conditions of actual service is given in the following description of a European experiment :

During the candidate's training period he is put through a test on a platform placed immediately in front of a large white screen. On this screen is projected a street scene taken by a cinema camera from the platform of a moving tramcar. The road winds; people cross in front; a taxi darts out from a side street; a man stops right in front of the car. The pupil manipulates levers and pedals as required, and his success is viewed as a measure of his fitness for the actual work (96).

The more usual form of miniature test is the test for motormen devised by Münsterberg. This apparatus is of special interest both because it was one of the first vocational-aptitude tests of any kind to be attempted and because it has served more or less as a model for subsequent tests aimed at the same general type of aptitude. The apparatus consisted of a black box covered by a piece of glass. Over the glass a black velvet belt was moved by means of a crank which was turned by the subject. In the belt was an opening or window. Beneath the glass was a large card. The opening in the belt was only large enough to expose a part of

the card at one time. On the card appeared two lines representing car tracks. At various distances from the tracks were digits in various colors representing various objects moving in various directions and at different speeds. As the window moved along over the glass one part of the card after another became visible. The task of the subject was to read off from the card as rapidly as possible the places of danger on the track which would be reached by all objects representing danger. The test was scored both as to the time required by the subject and his number of errors.

We turn now to the contrasted test type — the so-called “tests of abstract traits.” Let us suppose that we are still concerned with motorman aptitude but now are attacking it by attempting to isolate and measure the various component factors which make up the aptitude. In the analysis of this aptitude it might plausibly be assumed that quick (short) reaction time would be a desirable trait in a motorman on the principle that the more quickly a man could put on the brakes at sight of impending danger the less likely his car would be to encounter accident. The measurement of reaction time is one of the oldest and best-known experiments of the psychological laboratory. Used in the ordinary way, the test would be distinctly of the abstract type.

A common form of the simple reaction-time experiment is as follows: The main piece of apparatus is a chronoscope, a kind of clock for making very fine time measurements. This is connected by suitable wiring to an electric battery, in the circuit of which is a telegraph key. The subject sits comfortably at a table, with his finger on the key ready to press. He is instructed that shortly after a “ready” signal he will hear a click. At this he is to press the key as quickly as possible. The readings of the chronoscope reveal the time elapsing between the click and the reaction of pressing the key. This period is called the “reaction time.” It was used

extensively in Europe during the late war as a means of detecting latent aviation ability, though apparently with little success.

As a general thing the greater the difference between the stimulus-response situation embodied in the test and that found in the occupation in question, the less likely is the test to have aptitude-prognostic value. Thus, in the reaction-time test given above, if the stimulus had been given in the form of a touch on the hand and the response required had been that of closing the eyes as quickly as possible, the chance of the test being useful in forecasting motorman aptitude would have been even less than if conducted as described. But if instead the stimulus had been visual, the likelihood of success would have been increased, because in the main the eye is the organ which receives the stimulus in the actual occupation. And if, in addition, the reaction required had been of the same general nature as that used in setting the brakes on an electric car instead of closing the eyes, the chance of securing a useful test would have been still greater. And if, lastly, the stimulus had been a complex visual presentation resembling that actually encountered by a motorman in his daily duties, the test would have become a true miniature test. It will readily be seen from the foregoing that the distinction between miniature tests and tests of isolated traits is really a matter of degree. There is no sharp line of demarcation between them.

Miniature tests are likely to require rather elaborate and highly specialized apparatus. As a rule such apparatus is likely to be expensive, bulky, and not readily portable. As a compensation for these defects, miniature tests have the virtue of gathering together within a single organic behavior unit a relatively large number of the (too often unknown) determiners of aptitude success, under conditions favorable for joint quantitative determination. Perhaps of equal

importance, these factors are likely to be present in the test behavior in about the same proportions as in the occupational activity itself. Perhaps this is why tests of the miniature type are so apt to yield satisfactory aptitude predictions. One more serious defect of the miniature type of test in its extreme form is its sensitiveness to the influence of training. This is because the great similarity between the occupational activity and the test activity readily permits habits formed in the former to function in the latter. But if the psychologist goes to the opposite extreme and employs abstract tests involving responses highly different from those of the occupation in question, the tests are likely to have no prognostic value whatever. The escape from this dilemma seems to be to seek tests somewhere between the two extremes but tending rather in the direction of the miniature. This last is especially advisable where the aptitude aimed at is largely a matter of gross motor coördination and response to the coarser forms of external stimulation, as contrasted with internal stimulus-response sequences which constitute what we call thinking.

APPARATUS *vs.* NON-APPARATUS TESTS

A fourth division of aptitude tests may be made on the basis of whether or not apparatus is used. Of all apparatus tests, the most elaborate tends to be the miniature test just considered. The only strictly non-apparatus tests are the purely verbal tests of the oral type. However, many tests which are essentially verbal employ some conventional apparatus in the form of pencil and paper. Even the most purely oral-verbal tests we have, the Binet-Simon, ordinarily employ pencil and paper also, but for the most part these are in the hands of the examiner rather than of the subject; i.e., the subject's responses are recorded in the examination booklet by the examiner instead of by the subject himself.

A well-known test of the pencil-and-paper variety is the number-group checking test of Woodworth and Wells. A somewhat reduced copy of this test form is reproduced below. The procedure of the test is as follows: The sub-

THE WOODWORTH-WELLS NUMBER-GROUP CHECKING TEST
(*Reduced type*)

983642	168379	694517	253914	745682	158923	729648
426357	372159	754936	297835	627519	786531	731469
654173	947386	589761	134852	146237	194526	936425
837162	691324	814536	326175	368792	549826	572194
458671	971648	479612	495683	784295	817243	916328
275148	318495	635728	596873	982563	431289	381647
513978	182765	615832	851279	498136	356719	412789
197584	563792	748315	861395	421856	973124	125437
918654	846975	453867	281463	213956	651274	526987
397841	961872	248691	574389	532416	723964	473519
872351	327984	437528	864712	825916	682543	534169
923871	632791	765429	235849	672834	295481	349257
867314	462758	486592	198537	871596	164985	247153
963458	981374	156843	259671	762491	983567	579361
345962	941258	182653	561487	435781	179428	731825
672389	346521	427163	281937	672539	985273	956142
312876	853926	587436	296851	784623	875126	513647
934612	739548	843216	215367	916483	294378	768914
954178	371629	529817	436978	123874	957641	682917
719325	294736	639187	286415	593182	297568	145389
594231	389254	196235	825749	461289	378652	672841
349716	427395	138962	268794	524617	358472	319546
714932	759431	382145	853624	714529	635819	237465
649752	718254	596743	862934	851763	329418	495867

ject is given a pencil and is instructed at a certain signal to pass along each row of number groups from left to right and draw a mark through every group that contains both a 2 and a 5, say. He is to do this as rapidly and accurately as possible. This test was found useful by Link in the selection of bullet inspectors.

The classical example of the pencil-and-paper test is, of course, the Army Alpha battery. Under the pressure of war necessities pencil-and-paper tests received an amount of attention and consequent development which otherwise would not have occurred in a period of many years. The result is that pencil-and-paper tests are at present in a distinctly more advanced state of development than are apparatus tests of the more mechanical forms.

A common example of the more mechanical type of aptitude test is the peg board. This test shows a number of variations. A simple and useful one is shown in Figure 12. It is a piece of wood $5'' \times 15'' \times \frac{3}{4}''$ thick. It is pierced by 24 holes $\frac{3}{32}''$ in diameter and $\frac{1}{4}''$ deep, arranged in two rows. In a depression at the top lie a number of round wooden pegs just small enough to be inserted into the holes without pressure. The subject may be instructed to pick the pegs from the depression with his preferred hand, one at a time, and stick them in the holes as rapidly as possible. A number of variations may be carried out. For example, the subject may be required to do the task first with the preferred hand as just described and then with the non-preferred hand, so as to secure an indication of the degree of ambidexterity. Other variations are to require the subject to insert the pegs in the holes with both hands simultaneously; to insert them with the preferred hand, the pegs being fed to it by the other hand; and, lastly, to take the pegs out of the holes one at a time and replace them in the receptacle.

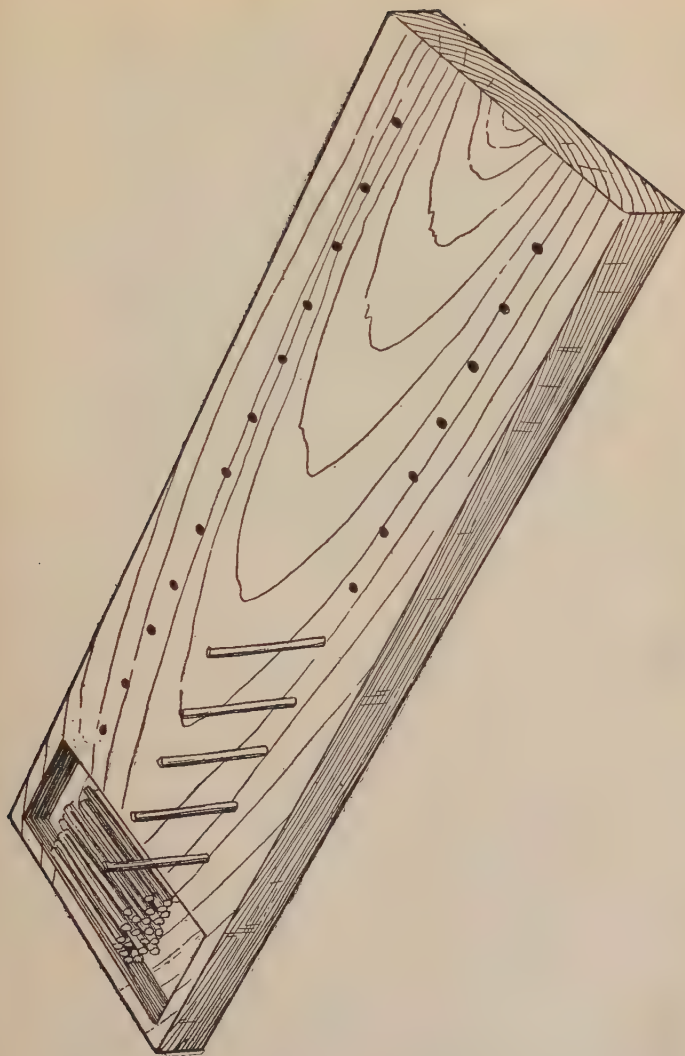


FIG. 12. The peg board.

With the peg board and most other apparatus tests, the score of the subject is to a certain extent a function of his voluntary effort. An example of a type of test not dependent upon the voluntary effort of the subject is the psycho-galvanic reflex. In this test two of the subject's fingers are placed in a circuit with a weak source of current and a very sensitive galvanometer. It is found that if a subject so placed is stimulated in a fearful or embarrassing manner, soon afterward the electrical resistance of his body decreases to such an extent that the galvanometer will show a marked deflection. It is believed that the phenomenon is due to the increased perspiration which normally follows such stimulations. To the moving part of the galvanometer there is usually attached a small mirror which reflects a beam of light. This beam may be thrown on a screen at some distance from the galvanometer and thus magnify the deflection very greatly. If it is thrown on a moving photographic film, there may be secured a permanent record of the psycho-galvanic response of the subject. Such a record is shown in Figure 13. The aptitude possibilities of this test have as yet been very little explored.

Among apparatus tests, the most striking contrast is that between the more mechanical apparatus on the one hand and the pencil-and-paper variety on the other. There is a field for both kinds of test, since obviously different kinds of behavior are sampled by each. It is probable, however, that a far greater variety of behavior may be sampled by the more mechanical type of apparatus than by the pencil-and-paper type. Owing to the fact that pencil-and-paper tests experienced a precocious development as a result of the war, it is likely that mechanical-apparatus tests have a much larger proportion of their possibilities yet to be realized.

An important aspect of all testing is the matter of ultimate cost. Apparatus — *good* apparatus — is likely to be ex-

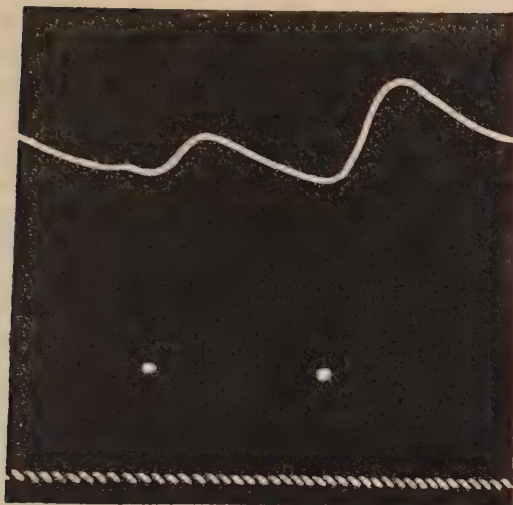


FIG. 13. Showing galvanic response to the threat of a pin prick.
(Wechsler, 100.)

pensive. Pencil-and-paper tests usually are regarded as inexpensive. This is true if only a half dozen or so subjects are to be tested. But it should be realized that the original cost of the apparatus is usually the only cost for an indefinitely large number of persons tested, whereas with the paper tests the continual need of buying new blanks is likely to constitute in the long run a more serious drain on the budget than the single original cost of the apparatus. This fact is too often overlooked by users of tests. Printed tests have, however, many very real advantages, as we shall presently see.

INDIVIDUAL *vs.* GROUP TESTS

A fifth grouping of aptitude tests may be made from the point of view of the number of subjects that may be tested by a single examiner at one time. This results in the familiar

division into *individual tests* on the one hand and *group tests* on the other. Historically, the first psychological tests to be used were individual tests. In this the psychologists were but following the traditions and practices of the early psychological laboratories. Group testing, of course, follows the general practice of giving course examinations long customary in the schools. At present, the best known of the individual tests is the Binet-Simon series. Probably the most widely known of the group tests is Army Alpha.

Wherever the response of the subject is spoken, the test must of necessity be individual. Obviously an examiner can neither record nor score the oral responses of more than one subject at a time. But quite apart from this consideration, it would be impossible to have two or more subjects giving oral responses within hearing of each other because the responses of the able would unavoidably give assistance to the less able, to say nothing of mutual distractions. Other tests which are likely to be of the individual type are tests which involve apparatus, especially if the latter be complex. The reason in this case is that the manipulation of a single apparatus, the taking of records, etc., are likely to require the entire attention of one examiner. There is also, of course, the cost of duplicating expensive apparatus which would be necessary if more than one person were to be tested simultaneously.

An example of an individual apparatus test is the psychogalvanic reflex described above (page 76). A second example is Henmon's test (30) for measuring the amount of involuntary jerking of the arm at the explosion of a pistol. The subject is seated in a chair, with arm extended at a convenient angle. He is given the apparatus to hold in his hand. The apparatus consists of a round metal cylinder of a size convenient for gripping, which terminates in a shallow metal cup about $2\frac{1}{2}$ " in diameter. Over the top of the cup is

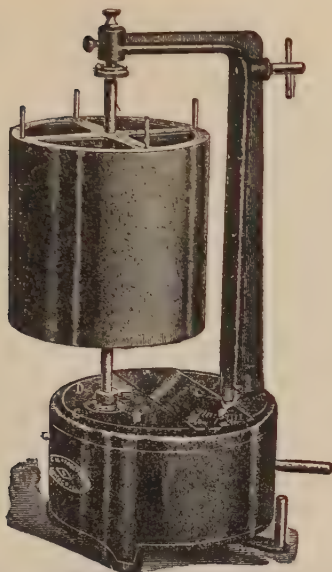


FIG. 14. A useful kymograph. The detachable drum of this instrument is covered with a glossy-surfaced paper which is coated by a film of soot by being revolved over a smoky gas flame. When placed in the machine the drum can be made to turn at a constant rate by a clockwork mechanism in the base. By a simple adjustment the rate may be varied within a rather wide range without stopping the drum. (Reproduced by permission of C. H. Stoelting Company.)

stretched some rubber dam so that the chamber within is air-tight, except for an opening to which is attached a rubber tube. On the middle of the stretched rubber membrane is attached a small metal weight. The apparatus is held in such a way that the weight of the piece of metal presses downward on the membrane. Now, if the subject makes any but the most gentle upward movements with his hand, the rubber dam will be depressed by the inertia of the weight resting on it, thus forcing air out through the rubber tube.

An opposite action will result if the instrument is abruptly lowered. The rubber tube connects with another rubber-covered cup similar to the first; only in place of the weight there is a support for a delicate lever, which moves with any movement of the membrane. This apparatus is called a "tambour." The end of this lever is placed in light contact with a drum covered with smoked paper and turned at a constant rate by clockwork. This is called a "kymograph" (Fig. 14). Now when a pistol is shot off near the subject, his hand is almost certain to jerk involuntarily. The disturbance thus set up in the air of the first cup is transmitted by the rubber tube to the second cup and is recorded on the smoked paper by the tracing lever. This tracing is made into a permanent record by being dipped into a thin solution of shellac. The score is the amplitude of the movement of the tracing lever as measured in millimeters or some other convenient unit.

A typical oral individual test is the auditory memory span for digits. In this test the examiner instructs his subject to listen closely while he (the examiner) reads off some digits. The digits are read off at the rate of one per second. At the end of each series the subject repeats the digits in exactly the order heard. That the test may be easy enough for the least gifted and difficult enough to take the measure of the most able, the test includes series of a wide range of lengths, as follows:

2 9 4

3 8 6

1 7 3

9 2 1 3

1 9 3 6

4 3 6 5

6 1 2 7 3

2 8 5 3 7

6 7 4 8 2

6 5 7 1 3 2

4 6 3 1 7 9

6 5 9 3 2 7

5 4 3 7 9 8 1

3 6 1 2 5 8 7

1 2 5 4 6 8 3

9 8 1 6 5 7 4 2

2 6 5 9 8 4 1 2

7 9 6 4 5 3 8 2

5 3 1 6 4 2 9 7 8

9 5 3 7 8 4 1 6 2

3 8 1 4 5 9 2 7 6

1 9 2 5 8 3 6 4 7 3

5 7 2 4 1 6 9 8 3 4

4 7 1 8 9 2 6 3 5 9

Each series reproduced without error receives a score of one. Most normal adults will go a certain distance down the list before making an error. The series missed rapidly become more frequent until the subject soon reaches a point beyond which he cannot succeed at all. After three or four successive failures the experiment is terminated. The number of series correctly reproduced is the score on the test.

Up to the time of the World War psychologists had a somewhat exaggerated impression of the inaccuracies of measure-

ment likely to result under group testing conditions. But when, during the war, they were faced with the necessity of testing enormous numbers of soldiers with an unprecedented speed and economy, they were forced to adopt group methods. It was then discovered that the inaccuracies of measurement in group testing were not so great as had been supposed. Moreover, numerous methods were discovered whereby the grosser forms of inaccuracy might be obviated. The result is that at present group testing is the dominant method in use. This dominance is due chiefly to its remarkable economy. By group methods a single examiner can often test from one to two hundred times as many subjects in a given time as by the individual method. It is recognized, however, that individual methods have a certain superiority as regards not only accuracy but range and completeness of observation of the subject's extra-test behavior.

THE CONVERSION OF INDIVIDUAL TESTS INTO GROUP TESTS

Many forms of individual oral psychological tests may very readily be converted into group pencil-and-paper tests. As a matter of fact, the majority of the group tests now in use were originally individual tests. As an example of the conversion of an individual test into a group test, we may consider the oral memory-span test described on page 81. This may be converted into a group test in the following manner: Simply have printed or mimeographed a form such as shown on page 83. Provide the subjects with these blanks and with medium hard pencils sharpened at both ends. The examiner then reads aloud the instructions at the top of the form, the subjects at the same time reading from their own copies. The examiner then proceeds to read off the digits shown on page 81, at the rate of one per second as in the individual form of the test. He pauses long enough after each series to permit the number to be written down if

TEST OF MEMORY FOR DIGITS

NAME.....DATE.....

DIRECTIONS. When the examiner has read off a series of numbers, write them down in the form below at once, in exactly the order read. Hold your pencil up until he has stopped reading and then (but not before) quickly write down the numbers. The first number is written in as it should be. Where you are not sure of a number, write down at once all you can remember.

294

remembered. The score, as in the individual form, is the number of series entirely correct. The reason for instructing the subjects to write down parts even of imperfectly remembered series is mainly to have an entry in each space so that any later and correct responses will be made in the right rectangles and thus facilitate scoring.

SOME TYPICAL PENCIL-AND-PAPER TESTS

The group test just described requires the examiner continuously to supply a vocal stimulus to direct the activity of the subjects. In most group pencil-and-paper tests, however, it is possible to supply both the directions and the material to be worked on by the subject by having them printed in detail on the blank which receives the graphic responses of the subject. This makes the test distinctly more fully "self-administering."

There are very many kinds of test material which may be presented by means of the pencil-and-paper type of group test. Some of these depend largely upon the results of previous learning, either casually acquired or systematically received in the schools. Others are dependent largely upon the facility of learning during the test itself. Still others depend upon visual discrimination or visual guidance of hand movements with the pencil. No test, however, is purely dependent upon any single type of activity, but all are complications of various activity factors. A dozen or so of the forms of test material, especially the more verbal ones characteristic of scholastic-aptitude tests, have become rather common. A goodly proportion of these were originated by Otis and Army Alpha, which Otis so largely influenced. A representative list, typical of the whole, follows. To secure a really adequate idea of the various test materials, however, the numerous forms should be examined in the test booklets actually used.¹

¹ See list of publishers given in footnote, page 307.

I. *Directions test.* A common form of the directions test was devised by Woodworth and Wells before the war. This test directs the subject to do certain things with his pencil. The difficulty does not lie in the manipulation of the pencil but in understanding the directions. In the Woodworth-Wells form the directions are printed so that the test becomes to a certain extent a test of reading ability. In Army Alpha No. 1 the directions are given orally, which largely eliminates the reading complication. The oral instructions for the first three items are given below in quotations. Beneath the instructions appear the forms to be marked by the subject:

1. "Look at the circles at 1. When I say 'go,' but not before, make a cross in the first circle and also a figure 1 in the third circle — Go! (Allow not over 5 seconds.)

2. "Look at 2, where the circles have numbers in them. When I say 'go' draw a line from Circle 1 to Circle 4 that will pass *above* Circle 2 and *below* Circle 3 — Go! (Allow not over 5 seconds.)

3. "Look at the square and triangle at 3. When I say 'go' make a cross in the space which is in the triangle but not in the square and also make a figure 1 in the space which is in the triangle and in the square — Go! (Allow not over 10 seconds.)"

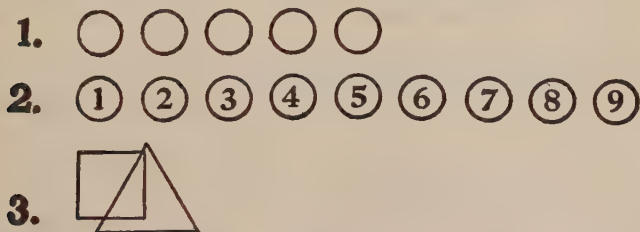


FIG. 15. Figures which are the basis of the first three items in the Army Alpha directions test. (Reduced 40 per cent.)

II. *Arithmetic.* A very common form of test involving verbal manipulation is to present a series of problems in ordinary arithmetic. Despite the very obvious dependence of this test upon scholastic opportunities, a recognized weak-

ness, it has survived in numerous aptitude tests. It appears as No. 2 of Army Alpha, three items of which are given to illustrate the type:

1. How many are 60 guns and 5 guns?.....*Answer* ()
2. If you save \$9 a month for 3 months, how much will you save?.....*Answer* ()
10. If it takes 4 men 3 days to dig a 120-foot drain, how many men are needed to dig it in half a day?.. *Answer* ()

III. *Best Answer.* This is classed by the army psychologists as a test of common sense. It requires a common-sense interpretation of information possessed by practically every normal adult. An example of this follows:

Cats are useful animals because

- ☐ they catch mice
- ☐ they are gentle
- ☐ they are afraid of dogs

Shoes are made of leather because

- ☒ it is tanned
- ☐ it is tough, pliable, and warm
- ☐ it can be blackened

The subject is directed to put a cross in the square before the words which complete the sentence in the most sensible manner.

IV. *Synonym-Antonym Test.* This test has appeared in a variety of forms and has proved of distinct value as a test of scholastic aptitude. It is essentially a test of meaning of words — partly a matter of knowledge and partly dependent on power of discriminating meanings. It appears as Test 4 of Army Alpha. The first and last two are given below. The subject is instructed to underscore the word *same* if the two words of a pair mean the same or nearly the same, and to underscore the word *opposite* if the words mean opposite or nearly the opposite:

1.	no — yes.....	same — <u>opposite</u>	1
2.	day — night.....	same — opposite	2
39.	supercilious — disdainful.....	same — opposite	39
40.	abstruse — recondite.....	same — opposite	40

V. *Disarranged Sentences.* This test is dependent upon the power rapidly to read isolated words and to arrange them into a meaningful sequence. As a test of whether this has been accomplished, the subject is required to state by underlining the appropriate word at the right, whether the resulting sentence is a *true* or *false* statement. This appears as Test 5 of Army Alpha. The first and last two items of this form are given below :

1.	oranges yellow are.....	<u>true</u> — false	1
2.	hear are with to ears.....	true — false	2
23.	feeling is of painful exaltation the.....	true — false	23
24.	begin a and apple scorn ant words with the ..	true — false	24

VI. *Number Series Extension.* This test involves the power of discovering the nature of the system upon which a number series has been built, and as an objective indication of whether or not this has been done successfully the subject is required to extend the series two more steps by writing the appropriate numbers in the blanks at the right of the series in question. This test appears as Test 6 of Army Alpha. The first and last two items of this test are here reproduced, the first being filled in as required :

3	4	5	6	7	8	9	10
8	7	6	5	4	3		
16	17	15	18	14	19		
3	6	8	16	18	36		

VII. *Analogies.* In some respects this test resembles that of the extension of number series. The subject must first discover the relation between the first two words and then underline the one of the four words in heavy type that

bears a similar relation to the third word. This appears as Test 7 of Army Alpha. The first and last two items of that test are reproduced below, the first item being worked as it should be.

1. shoe — foot : hat — kitten head knife penny
2. pup — dog : lamb — red door sheep book
39. book — writer : statue — sculptor liberty picture state
40. wound — pain : health — sickness disease exhilaration doctor

Dr. Lewis Anderson has devised an extremely ingenious test on this principle by using geometrical figures. Six items of this test are reproduced in Figure 16, the answer to the first item being marked as it should be (by underlining).

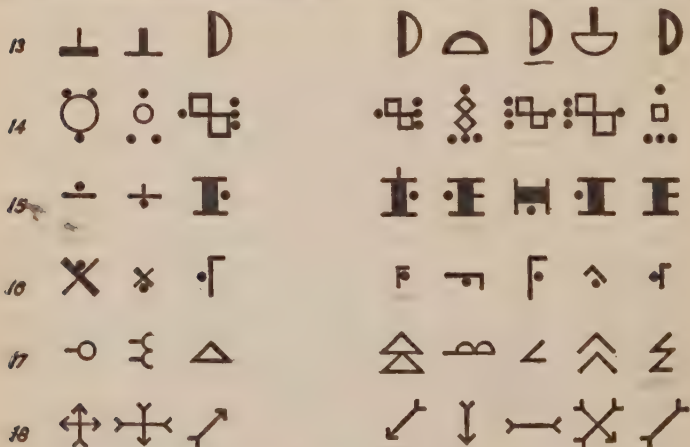


FIG. 16. Part of the Anderson mixed-relations test. (Reproduced by permission of the author. Reduced $\frac{1}{4}$.)

VIII. *General Information.* This test depends upon the information possessed by the subject. It is doubtless to a considerable extent influenced by school opportunities, but is also largely influenced by the alertness of the subject in picking up miscellaneous bits of information from the

ordinary contacts of life. This material makes up Test 8 of Army Alpha. The first and last pairs of items of that test are reproduced below. The first is marked as it should be.

1. The apple grows on a shrub vine bush <u>tree</u>	1
2. Five hundred is played with rackets pins cards dice	2
39. Napoleon defeated the Austrians at Friedland Wagram Waterloo Leipzig	39
40. An irregular four-sided figure is called a scholium tri- angle trapezium pentagon.....	40

IX. *Proverbs*. Another test which has proved of considerable value in scholastic-aptitude tests is the interpretation of proverbs. A series of proverbs more or less cryptic in expression is given, after which is given another series with similar meanings but differently arranged, or perhaps a series of straightforward statements of meaning:

- (1) All is not gold that glitters.
- (2) Actions speak louder than words.
- (3) Where there is much smoke there must be some fire.
- () Good deeds performed are more dependable than promises.
- () Where there is much indirect indication of evil there is likely to be a certain amount of real evil.
- (1) Appearances are often deceptive.
- () A friend in need is a friend indeed.

The subject is required to place the number preceding each proverb before the sentence in the second series which expresses the corresponding idea. The expression equivalent to the first proverb is already marked as it should be.

X. *Sentence Completion*. This test consists of a series of printed sentences in which significant words are omitted and are replaced by blank spaces. The subject is required to write in the omitted words. Ebbinghaus, the German psychologist, many years ago expressed the opinion, from theoretical considerations, that this type of test would be valuable. Subsequent experience with it has gone far to

substantiate his prediction. The test suffers somewhat from the fact that often a number of different responses are more or less correct. This makes the test rather more difficult to score than most group pencil-and-paper tests. Trabue has standardized a large number of such tests of graded difficulty. The following sentences illustrate the type :

The boy _____ the stone across the river.

Fire and water _____ mix.

If a man _____ on his head _____ probably break his _____.

XI. Cause and Effect. In this test the cause is given as a single word and the subject is required to select the effect (a noun) from four other words, underlining the one chosen. This appears in the Miller Mental Ability Test, published by World Book Company. The author gives as examples :

1. FIRE (hot, house, damage, melt)
2. WATER (cold, flood, thirst, transparent)
3. ELECTRICITY (wire, snow, light, hot)

XII. Most Closely Related Words. This is an interesting form of verbal group test used by the College Entrance Examination Board as a unit in a battery for predicting collegiate academic aptitude. The first three of the items of their list will serve to illustrate the principle of the test. The first item is marked as it should be.

doll ₁	ring ₂	flower ₃	drum ₄	top ₅	shoe ₆		1	4	5
bean ₁	carrot ₂	potato ₃	beet ₄	lettuce ₅	cabbage ₆		—	—	—
ace ₁	hearts ₂	spades ₃	cards ₄	trumps ₅	diamonds ₆		—	—	—

XIII. The Porteus Maze Test. This test consists of a maze printed on a sheet of paper. The starting point is marked by *S*. The subject is given a pencil and told to trace a pathway from *S* to the outside as quickly as possible but without crossing any lines. One of the more complicated forms of this test is shown in Figure 17. The irregular line from *S* to *O* is

the record of an actual subject. False moves were made at *X* and at *Y* which necessitated retracing. This test was

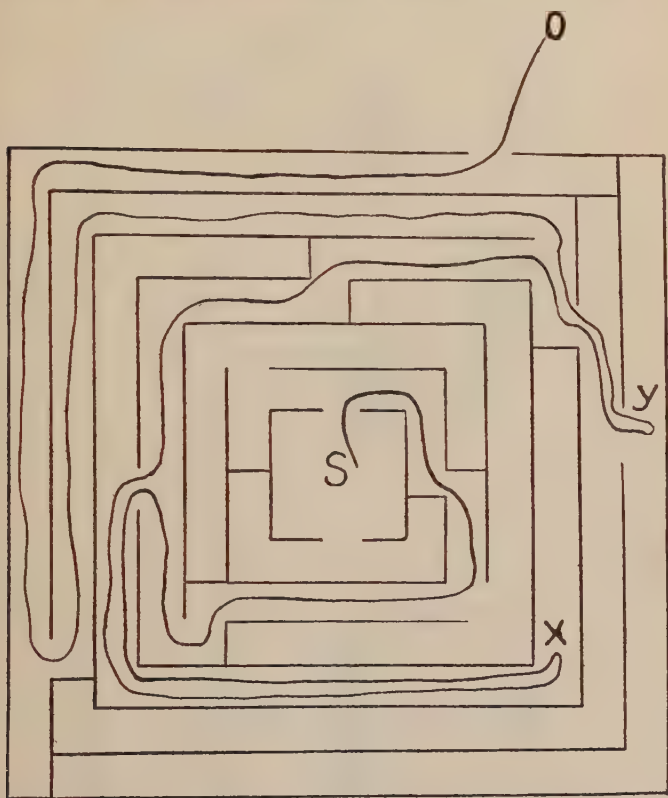


FIG. 17. Showing the record of a subject on the Porteus pencil maze test. (Reproduced by permission of C. H. Stoelting Company. Reduced $\frac{1}{4}$.)

designed to sample the tendency to look forward and consider the consequences of an action before executing the act. It appears, therefore, to be intended as a test of prudence.

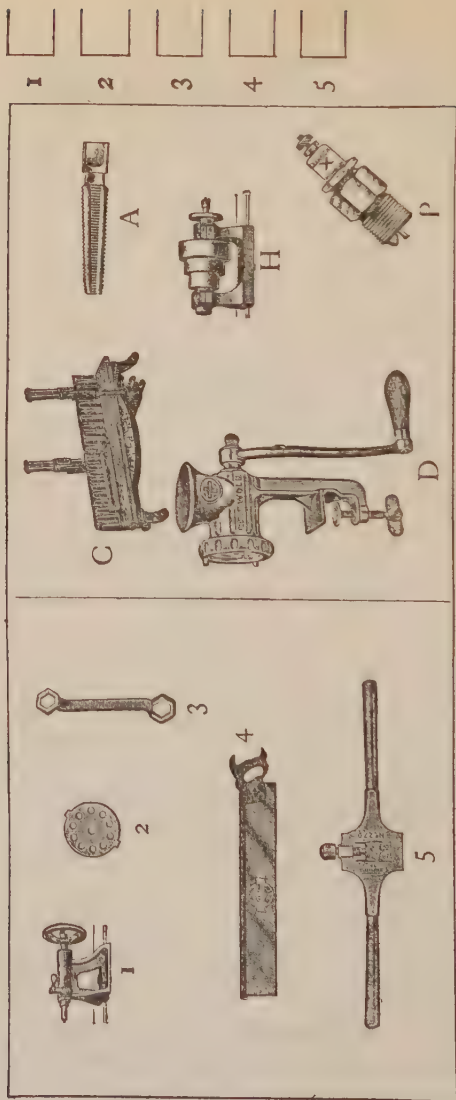


Fig. 18. Part of the Stenquist picture test for mechanical aptitude. (Published by World Book Company.)

XIV. *Mechanical Information.* As a last example of a pencil-and-paper group test may be mentioned the mechanical-aptitude test of Stenquist, based on mechanical information. There are several different parts to this test (Fig. 18). Each of the different parts is based upon a series of pictures of various mechanical devices more or less familiar to persons interested in such matters. One of the characteristic parts of this test is here reproduced.

EXAMPLES OF GROUP TESTS INVOLVING APPARATUS

The possibility of utilizing apparatus to present auditory stimuli for purposes of group testing has been employed very cleverly by Seashore in his tests for musical aptitude. He has combined in an extremely effective manner (*a*) a general oral instruction, (*b*) phonographic methods of producing musical and other tones, and (*c*) the recording of the responses by the subjects on simple blank forms with a pencil. The general method is to present phonographically the tones or groups of tones involved in a given test in pairs, the subject marking down on the blank by means of a specified letter his reaction. For example, in pitch discrimination, if the second of the pair sounds higher to him, he marks down an *H*; if it sounds lower, he marks down an *L*. Similarly in the test of consonance. If the second combination sounds worse, — i.e., less harmonious than the first, — he marks down a *W*; if better, — i.e., more harmonious, — a *B*. In this way Seashore has developed five tests of musical hearing. Besides the tests of pitch and consonance already mentioned, there are tests of time, intensity, and tonal memory. It is highly probable that with the continued development of phonographic methods of sound reproduction, this method may be utilized for tests other than musical hearing. The synchronization of sound reproduction with motion pictures is too new as yet to be evaluated as a testing

medium, but it probably has great possibilities. At their best, such methods have the great advantage of uniformity of presentation of the test stimuli. They have the defect of being non-responsive to emergencies which may arise in the testing situation.

It has already been stated above that apparatus tests tend strongly to be individual rather than group in their mode of administration. There is reason to believe, however, that the future will see a considerable adaptation of apparatus to group methods. The chief difficulties to be met are three in number: (1) The cost of duplicating the apparatus, (2) the difficulty of making sure that the subject understands how to operate it, and (3) the problem of securing an accurate score of the subject's behavior. These problems are probably sufficiently formidable to limit apparatus group testing to relatively small groups — perhaps to groups of 20 or 25 as a maximum.

As an illustration of the solution of these difficulties in an important aptitude test may be mentioned the adaptation of the Stenquist mechanical-assembly test to group conditions. This test consists essentially of ten small mechanical devices, such as a paper clip, a bicycle bell, a simple lock, etc., which are presented with the parts disassembled. The task of the subject is to assemble as many of them as possible and as well as possible in a specified time. The various objects, together with the case in which they are presented, are shown in Figure 19.

The Stenquist test materials are sufficiently inexpensive to make it practicable to duplicate them for group-testing purposes. The problem of making sure that the subject knows how to operate the apparatus does not exist in this case, because that is the essence of the test. The final problem of scoring is solved by doing the scoring after the subjects have finished. This is possible because the scoring

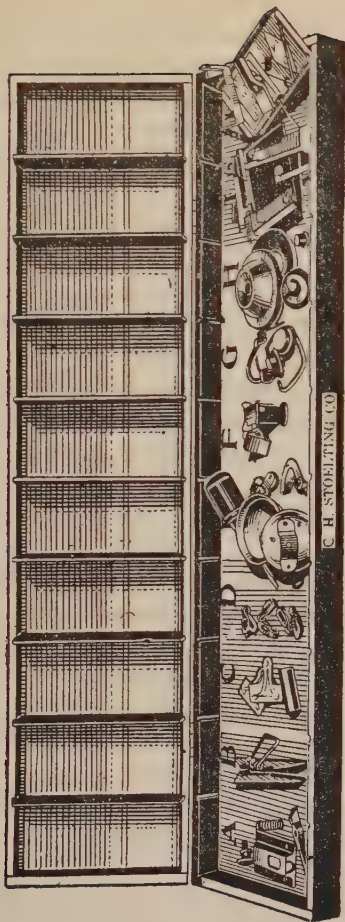


FIG. 19. The Stenquist assembling test of general mechanical ability. (Reproduced by permission of C. H. Stoelting Company.)

is done on the basis of the state of the test objects as left in the box by the subjects. It may be added that in the group form of the Stenquist assembly test the subjects are prevented from watching each other, by means of small screens placed between the various individuals.

The mechanical-assembly test just described presents especially favorable conditions for group scoring. Most apparatus is not so favorable. It is probable, however, that as time goes on and psychologists realize the advantage of sampling a wider range of behavior than that afforded by simple pencil-and-paper methods, special *self-recording apparatus* tests will be developed for testing a great variety of behavior. A promising beginning in this direction has been made by Sidney L. Pressey. Pressey's apparatus consists essentially of a drum upon which may be placed printed slips of paper containing any kind of multiple-choice material desired. The drum exposes the material at a window, one item at a time. The subject responds by pressing keys having numbers corresponding to the numbers of the items in the multiple-choice material. Whenever the subject presses a key, the drum moves up so as to expose a new item. If the right key was pressed, this is recorded as a success on a special automatic counter; if the wrong key was pressed, the fact is recorded as an error on a second counter. Thus at the end of the test the examiner has only to glance at the counters to know the exact score. This machine may be set so as to do numerous other testing duties. It may even be made to reward the subject at the conclusion of a difficult task by presenting him with a tasty bit of candy! It is said that even adult subjects rarely refuse the proffered reward.

SPEED vs. POWER TESTS

A sixth grouping of aptitude tests of some significance is based upon the contrast between *speed* and *power* in test

performance. There is a very widespread popular belief that some individuals may be quick but superficial and that other individuals may be slow but sure in their responses. In other words, it is sometimes believed that the ability to do rapidly what one is able to do, does not indicate any special capacity to perform the more difficult feats contained in a test. Indeed, popular psychology, fixing its attention as it frequently does upon the exceptional, often seems to assume a negative correlation between speed and power in test, as well as in other, behavior. In accordance with this distinction, we find some tests so designed that the various items are fairly uniform in difficulty, as in the tapping test, the score being dependent primarily upon speed. On the other hand, there are tests which have the items arranged in an order of uniformly increasing difficulty but with no time limit, the point of the test being to see how difficult tasks the subject can achieve when given all the time he needs. An example of such a test is the Psychological Examination designed for high-school graduates and college freshmen, which has been published by the American Council on Education. In this battery the time allowed is so great that practically all subjects are able to finish. By far the greater number of tests in current use, especially the pencil-and-paper group tests, combine the two principles by having the test items arranged in the order of difficulty, but also with a rigid time limit. These latter tests are often (though rather misleadingly) called speed tests also.

Despite the very general belief to the contrary, speed and power in test performance are positively correlated, and rather highly. Indeed, the correlation is so high that some writers have concluded that in the Army Alpha type of test speed may, for all practical purposes, be taken as an indication of power. A typical investigation of this kind is that of Ruch and Koerth (72). They gave Army Alpha to 122

college freshmen and scored separately (a) what was accomplished in the regular time allowance, (b) the amount accomplished when double time was allowed, and (c) the amount accomplished when unlimited time was allowed. It was found that single time correlated $+ .966$ with double time and $+ .945$ with unlimited time. The latter coefficient was regarded as the correlation between speed and power. The interpretation of this coefficient would have been clearer if the speed test had been made up entirely of items easy enough so that all subjects could have performed almost all of them correctly. Nevertheless, the indication is very clear, not only from the above investigation but from others, that speed and power are decidedly and positively correlated.

But even though speed and power in test behavior are correlated as much as $+ .90$, it by no means follows that the two are identical. Under the conditions assumed it would be possible to estimate the one, knowing the other, with an efficiency of only 56 per cent. It is altogether probable that there are genuine cases where distinct power may be associated with deliberateness of reaction, and vice versa. It would be distinctly unsafe to assume that this tendency would be non-significant for all aptitudes.

While it is probable that the present practice is stressing a little unduly the speed factor in testing, it must be noted that there is a certain difficulty in the administration of group test batteries of the semi-power type where the time limit is so long as to permit nearly all to finish. This arises from the fact that some individuals finish a given test unit long before the slowest ones and many of these more speedy individuals become restless while waiting for the time to begin the next unit. In such cases it is difficult to prevent a certain amount of disorder and confusion with a large group. This difficulty would not be encountered if the material were arranged in the omnibus form, or if the subject were

permitted to begin the next unit as soon as he finished the preceding one.

TIME-LIMIT *vs.* WORK-LIMIT TESTS

Speed tests may be divided on the basis of the mode of timing into *time-limit* tests and *work-limit* tests. A simple example will make this distinction clear. Take the number-group checking test shown on page 73. Now if a subject is told to check in this test form as rapidly as possible every group which contains both a 2 and a 5, and if the time required to do this is taken, say, with a stop watch, the test would be known as a work-limit test. This is because the only limit set by the test is the amount of work to be done. The score is in terms of minutes and seconds. On the other hand, if the subject is instructed to check as many groups as possible containing both a 2 and a 5 in a period of 2 minutes, the test would be a time-limit test. This is because the limit set by the test in this case is the length of time to be spent. By this latter method of timing, the score will be the number of groups correctly canceled. By the work-limit method the larger the score (the more time consumed) the *worse* the performance, whereas by the time-limit method the larger the score (the more work accomplished) the *better* the performance. From this it is evident that scores received from the same test activity by the two methods will correlate negatively.

Many tests, like the number-group checking test given above, may be scored by either method without much distinction. Of the two methods, however, the time-limit method has a decidedly wider usefulness. Indeed, group testing is confined exclusively to the time-limit method. On the other hand, practically all individual tests may be given by the work-limit method, though the most of them may also be given quite as readily by the time-limit method.

The only tests which require the work-limit method are those rare ones on which the test behavior is largely an organic whole or at least is not made up of comparable part-activities which can serve as the units of test accomplishment. Mechanical-assembly tests (page 94) tend to be of this type, though complications arise with such tests where a subject is quite unable to perform the task correctly no matter how much time is given. Another example of a somewhat similar test situation is in the cube-assembly test. In this test a 3-inch cube, painted on the outside, is cut into 27 1-inch cubes and the subject is required to assemble it so that the painted surfaces will all be on the outside.

TESTS OF VARIOUS STIMULUS-RESPONSE MECHANISMS DIFFERENTIATED

An eighth class of aptitude tests may be made on the basis of the familiar stimulus-response category. Thus we have tests of stimulus-receiving mechanism; i.e., the efficiency of the sense organs. These tests are ordinarily called *sensory* tests. Similarly, there may be tests of the efficiency of the motor-responding mechanism. These are usually called *motor* tests. Lastly, there may be tests of the efficiency of the little-known mechanisms which mediate between patterns of stimulation, on the one hand, and patterns of response on the other. These tests, depending somewhat upon their supposed degree of complexity or subtlety, are usually called *mental*.

It must be observed at the outset that all tests involve all three phases of bodily activity. Consider a typical test of motor ability, such as strength of grip as measured by the hand dynamometer. (See Figure 20.) The subject must have intact sense organs (auditory or visual, say) to receive the stimulus pattern which constitutes the examiner's instructions, an intact central nervous system to transmit

the impulse to the particular muscles involved, and intact muscles to execute the response. Similarly, in a typical

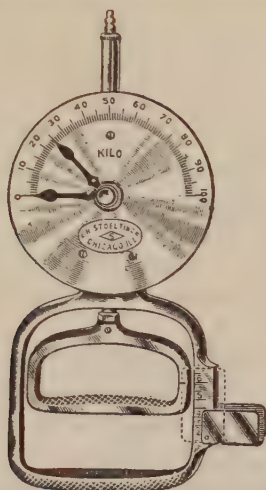


FIG. 20. The Smedley dynamometer. (Reproduced by permission of C. H. Stoelting Company.)

sensory test like that for color-blindness there must be, in addition to the eyes, an intact central nervous system and intact speech or other muscles by the action of which there is conveyed to the examiner an indication of how well the sense organ is functioning. And where a test is designed to measure the efficiency of the inner mediating mechanism in a typical mental performance, such as problem solving, there must be not only an intact mediating apparatus, but also an intact sensory equipment for receiving the examiner's instructions and an intact motor (usually speech) mechanism for indicating to the examiner the nature of the solution.

Since tests designed to measure the efficiency of any one of the three different types of mechanism always involve the action of the other two types, the question naturally

arises how we may be sure in any particular case which of the three mechanisms is really determining the response. The answer is that in more than a few cases there is genuine uncertainty. Indeed, it is probable that in all tests there is a trace of the influence of all three mechanisms. A case clearly illustrating this compositeness of the determining factors is the familiar cancellation or A-test, shown on the following page. It is likely that the performance on this test is dependent jointly upon the acuity of vision, the precision of ocular fixation which involves the action of the eye muscles, the efficiency of the central mediating apparatus, and the ability of the fingers to execute rapid and precise strokes with a pencil.

Despite the seemingly hopeless blending of sensory, motor, and mental efficiency in all forms of test behavior, the method whereby the efficiency of each is measured in isolation sufficiently well for practical purposes is very simple. It consists merely in so setting the task for the subject that the part to be performed by the two divisions of the reacting mechanism not to be tested is so easy that it can be performed perfectly by all ordinary subjects. In this connection we may recur to the test for color-blindness. The requirements of this test with respect to sensing and understanding the directions, on the one hand, and as regards the strength of muscles involved in making the response, on the other hand, are so slight that no person, unless downright feeble-minded or about to collapse from physical weakness, would fail in the test except through defective vision. If the color-blind test be properly designed, no amount of problem-solving ingenuity or physical strength should be able to compensate to any appreciable degree for a defect of vision.

THE A-TEST

(A test of rapidity of canceling A's)

ARYFRNJLVAAABAHYKKQAKPALEWNAIPAAEHANAO
 AZQUWBIJXAGLONDSEMRJYQAFVUYADOAYADCWA
 AFRAAKUATAAAWCFZAAUDFEHNSLBAXBACZJOVAW
 IUMUYZXAJFDGKTOFFSAOCLAMVYVASDZAJAIATBAC
 AGYHUZSHRGQJTNAPJRWOPQDCMAUEEIMNXDGTAQ
 AMUIXATOPOGCTNVAGXKAARCAADAHQDBAFXLDWYI
 RWXBGXBIUQADSJNMPAGAVAWIYZMXMNZTEAAAZA
 MAHHMGKPIAAWCGNSWALTQRYAKCGDXKSTAVUXX
 KLPVAOYAASQLRVXISOADSQYTVILPAOVJUAENNJKP
 SPOULVRBJQUANAAJHFBIGQHVAVENAICIPRENTLA
 AAWBGNWSQRXOXGOCHPSQUQWBLARRAZAZIEMDT
 APKNKBCAJZPBSLRJHHEASFFTNAVWBWZAEAAALEWA
 HCRVASTYWALAUVGFKHEADHOFABQKCTAFAYNCWHA

Name..... Date.....

Test No..... Score.....

TESTS OF SENSORY EFFICIENCY

Fairly adequate tests have for some time been available for the more important senses, particularly for sight and hearing. An excellent account of the more standard sensory tests is given in Whipple's *Mental and Physical Tests, Simpler Processes*. This work should be consulted by any one interested in such tests. In the volume mentioned Whipple gives with many illustrations a full account of the following sensory tests:

Visual acuity

Balance and control of eye-muscles

Color-blindness

Discrimination of brightness

Auditory acuity

Discrimination of lifted weights

Discrimination of pressure

Sensitivity to pain

Discrimination of dual cutaneous impressions

Discrimination of pitch

As an example of a useful sensory test we may take the Holmgren wool test for color-blindness. The most common form of color-blindness is defect in vision for reds and greens. For a long time the importance of a test to detect this defect has been recognized and several different methods have been developed for the purpose. In addition to railway and marine service in which employees must see colored signals, many other vocations also require good color vision. One of the earliest tests is that of Holmgren. It consists of a considerable number of small bundles of yarn of various colors. In addition there are three larger bundles: a pale green, a rose, and a red. The test is conducted by first handing the subject the large green skein and asking him to choose from the small ones all those which are the same general color.

The color-blind are likely to show much hesitation and will probably choose along with some green skeins, also some grays, browns, and even pinks or reds. The remaining large skeins are used in a similar manner for purposes of confirmation of the test as shown by the green skein.

TESTS OF MOTOR EFFICIENCY

A number of useful tests have been devised and standardized for purposes of measuring physical and motor capacities. An important division of these tests concerns the measurement of gross muscular strength. Fairly satisfactory tests have been devised to measure the strength of hand grip, of the legs, of the back, and of various other muscle combinations. For the most part the apparatus employed involves a coil spring of known strength, which the muscles to be tested put under tension much on the principle of a spring scale. An account of the more important tests of muscular strength is given by Whipple.

Perhaps the most common and useful single instrument for measuring muscular strength is the hand dynamometer. An excellent form of this instrument is that designed by Smedley and known by the designer's name. An improved form of this is shown in Figure 20, on page 101.

A second division of motor tests, differing considerably from those for muscular strength, includes tests for motor coördination, agility, and skill. Among such tests which have become somewhat standardized may be mentioned the various tests employing the peg board (Fig. 12), the tapping test, the steadiness test, the tracing test, the aiming test, and the three-hole test. An interesting form of the three-hole test is one devised by F. G. Mueller, shown in Figure 21. In this test the subject is directed to strike the bottom of each hole with the stylus as rapidly as possible without striking the edges. The apparatus is so constructed that every time the

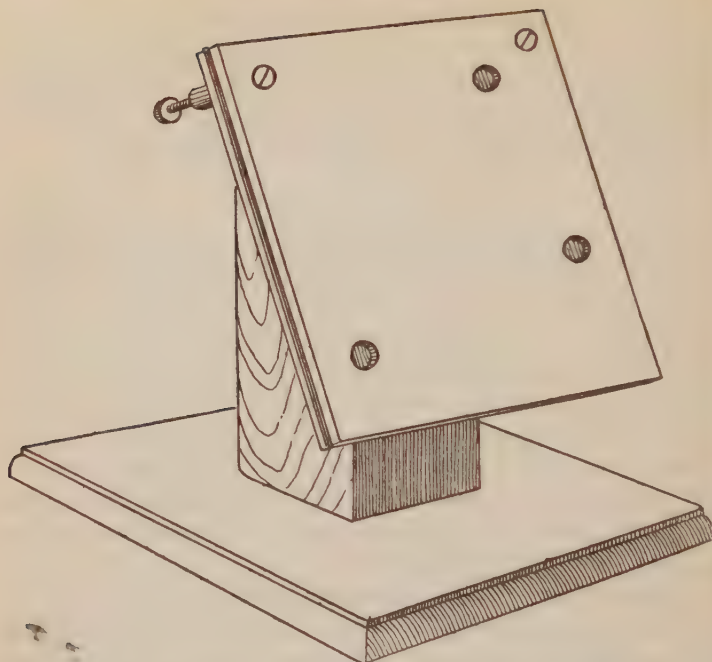


FIG. 21. Mueller's three-hole test. The stylus is not shown.

stylus strikes the plate at the bottom of a hole the stroke is registered on an electric counter. The subject may be directed to insert the stylus in the holes one after the other in a clockwise direction for a period of one or two minutes. He is warned to avoid hitting the edges as far as possible. In case the edge of a hole is struck by the stylus, this is recorded on a second electric counter as an error.

TESTS OF "MENTAL" EFFICIENCY

By the term *mental* is here meant those central processes which mediate between the external stimulus pattern on the one hand and the explicit response pattern on the other.

The greatest variety and complexity of tests are found here. These are the tests which are usually meant by the expression "psychological" tests, in the more narrow sense. Here fall the various tests involving learning, memory, and various forms of problem solving. Here we find such tests as description of an object, fidelity of report, controlled association, uncontrolled association, analogies, substitution, memory span, linguistic invention, word building, sentence completion, interpretation of fables, matching proverbs, and so on.

As a final example of this great family of tests there may be given the Woodworth-Wells substitution test. In this test the subject is directed to place in each figure encountered in moving across the lines from left to right, the number found in the corresponding figure at the top of the page. This test seems to be based largely upon the capacity to learn, though it is probably somewhat dependent upon rapid and accurate visual fixation. This may serve to emphasize once more the fact already remarked, that all tests probably are dependent more or less upon all three parts of the reacting mechanism. In the great majority of tests the three are inextricably mixed.

TESTS OF CHARACTER AND TEMPERAMENT

The man on the street and the man of affairs have always been interested in character. The early aptitude psychologists, on the other hand, were engrossed by the symbolic, verbal, or intellectual phases of behavior largely to the neglect of the more general and elusive tendencies to actions just mentioned. The layman, lacking many of the inhibitions of the trained psychologist, consequently has exploited this important field almost exclusively. Within the last few years, however, the psychologists have begun to realize that the ordinary tests of ability were failing to tap extremely important determiners of aptitude, and recently a vast

amount of energy and ingenuity has been devoted to the devising of tests intended to measure in one way or another a great variety of character traits. The task, as was anticipated, has proved a difficult one. As yet no very great success has been attained, but the future holds considerable promise.

One of the most enterprising of the investigators in this field is June E. Downey. She took her point of departure from the activity of handwriting. Basing her tests upon ingenious conjectures as to the nature of various aspects of action such as speed of decision, freedom from load, motor inhibitions, motor impulsion, self-confidence, and finality of judgment, she attempted to evoke generalized reactions on the basis of handwriting. For example, "freedom from load" in general was thought to be indicated by the difference between the ordinary rate of writing and the rate under command to write at top speed. As a second example, motor inhibition was thought to be revealed by the slowness with which writing could be done when the task was to write *just* as slowly as possible. This latter test is described in some detail as employed in an aptitude investigation (pages 361 ff.), where it is called the "slow movement" test. While of considerable value as a pioneering investigation, the Downey Will-Temperament Tests have shown a fairly uniform lack of validity or practical value, despite very widespread attempts to use them.

Numerous other schemes have been experimented with. Fernald and others have sought to secure an index of the normality of the order of ranking of a series of descriptions of offenses by a given person as compared with the order as arranged by the average normal adult. Pressey in his X-O tests has devised an interesting series of group tests which are designed to indicate deviations from normality in various directions with respect to such matters as suspicion, disgusts,

moral disapproval, etc. Myerson has prepared a suggestive test in the form of an unfinished story. The story is made up of a series of sentences, the first parts of which are given, and the subject is asked to complete each sentence by several alternatives presented for his choice. The temperament of the subject is supposed to be revealed by the character of his choices. Fernald has a test designed to reveal will power, or persistence of an action despite the presence of stimuli calculated to inhibit it. He asks the subject to stand with his heels one fourth of an inch off the floor as long as he is able. It is assumed that the longer a person is able to resist the fatigue and pain which result, the stronger his will.

One of the most promising series of the tests designed to measure character traits has been worked out by Voelker (97). In this study the purpose was to discover whether Boy Scout training in ideals of conduct really functioned in practical situations. To this end he developed an ingenious series of tests involving practical activities in everyday life. One of the most suggestive of these will be described.

The Voelker Profile Honesty Test. "Can a subject be trusted not to peep when he is placed on his honor to keep his eyes closed?" In this test Voelker uses a modified form of the Pintner profile test. This test is based on a profile of a man's head cut out of a thin board. Several of the prominent features of the profile, such as the nose, the ear, etc., are made up of separate pieces. The ear in particular is in four pieces and painted on the upper side to resemble a real ear. After the boy who is being tested has been made familiar, by a little practice, with putting the profile together, he is told to repeat the test *with his eyes closed*. The experiment is repeated three times in this way, the subject being instructed to open his eyes when he thinks the profile is perfect. Now the four blocks making up the ear part of the board

will fit as well one side up as the other, and the difference is not distinguishable by touch alone. Therefore the chance of all four blocks being placed right side up without vision is 1 in 16, and the chance of this taking place three times in succession is only 1 in over 4000, or practically nothing. Accordingly, if a boy makes three perfect successes it is practically certain that he has cheated.

CHAPTER FOUR

ANATOMICAL AND OTHER ALLEGED SIGNS OF APTITUDE

IN the preceding chapter we have considered various types of aptitude tests. We have now the task of examining a number of approaches to aptitude prognosis which, while distinctly allied to tests, should probably be regarded more as *signs*.

APTITUDE TESTS *vs.* APTITUDE SIGNS

An aptitude test (page 4) is ordinarily thought of as a sample of a relatively simple component of the aptitude behavior which it is desired to prognosticate. A sign, on the other hand, should not be regarded as a component of aptitude behavior but merely as a more or less reliable indicator of it. The nature of the connection or relation between the sign and the aptitude may or may not be known. From a certain narrow point of view it may be said to be irrelevant for purposes of prognosis, whether the connection be known or not. At any rate, the critical consideration is that the relation between the sign and the aptitude be known to exist and to be reasonably close.

An example will clarify this distinction. A man may be requested to exert his utmost power in manipulating some form of spring dynamometer. The sample of his behavior thus obtained should be useful in forecasting the aptitude of the man for doing a heavy type of labor involving substantially the same muscles. This would clearly be a *test* of strength. But suppose we had measured, instead, the girth of the man's arm over the biceps when the arm is bent rigidly at the elbow, or even taken the man's height, weight, and girth of neck. We might secure from these data also an

indication of a man's ability to perform some forms of heavy work. These latter measurements would be *signs* of strength.

The nature of the connection between the sign and the aptitude in this case, while by no means perfect, is sufficiently obvious. But when we pass from using height and weight as signs of aptitude for heavy labor to using them as signs of aptitude for becoming an executive of a corporation, the relation, while possibly real, is by no means so obvious as to the nature of its connection.

Some of the signs which we shall be considering in the following pages have been the subject of more or less violent controversy. Among these may be mentioned physiognomy, phrenology, chiromnomy, and graphology. On the other hand, there are various alleged somatic types of recent discovery which purport to be based on the activity of one or another of the endocrine glands or their combination. It has been asserted that these somatic types are accompanied by fairly well-defined psychological or behavior types. Lastly there will be considered certain possibilities, but little explored as yet, of discovering relations between physiological processes, conditions, secretions, basal metabolism, chemical constitution of the blood or saliva, etc., on the one hand, and aptitudes on the other. It is our purpose to survey the scientific evidence available in the several cases in as impartial a manner as possible.

TYPICAL CLAIMS REGARDING ANATOMICAL SIGNS

Perhaps the most ancient of all the beliefs regarding the signs of mental constitution relate to physiognomy and phrenology. These beliefs, like most others of the same general class, seem originally to have been supported largely by superficial analogies. Thus we find the following observation attributed to Aristotle:

Those who have a large head are sagacious — like dogs; those who have a small head are stupid — like asses; those who have no shame are like birds with curved claws.

Throughout the centuries the belief in physiognomy as indicative of mental traits has persisted with astonishing tenacity. From time to time individual enthusiasts have come forward with elaborate systems by which the character of the individual was supposed to be revealed by his face. In addition to the superficial analogies already mentioned, the evidence offered has been, in the main, chance observations of striking facial characteristics of famous persons. As an example of the claims of modern physiognomists, the following, taken from a work which has had considerable vogue in recent years, will serve as well as any :

The significance of the pure convex type (of face) is energy, both mental and physical. Superabundance of energy makes the extreme convex keen, alert, quick, eager, aggressive, impatient, positive, and penetrating. The tendencies indicated by his convex mouth will cause him to speak frankly and at times even sharply and fiercely without much regard for tact or diplomacy. . . . The pure concave, as might be expected, is the exact opposite, so far as the indications of form are concerned, of the pure convex. The keynote of his character is *mildness*. . . . He is slow of thought, slow of action, patient in disposition, plodding. . . . The convex is also, in the majority of cases, a blond. The combination of hopeful, optimistic, restless, organizing, creating, domineering characteristics of the blond with the quick, alert, practical, aggressive qualities of the convex make this type distinctly the type of action.¹

Owing largely to the late development of experimental psychology, though no doubt in part also to the somewhat elusive and intangible nature of character traits, it has been only within a decade or so that any really scientific evidence

¹ Blackford, Katherine, and Newcomb, Arthur, *The Job, the Man, the Boss*, page 154. Quoted by permission of Doubleday, Page & Co.

regarding the claims of physiognomy and other alleged anatomical signs of character and aptitude have appeared. Within the last few years several studies have been made which are gradually enabling us to sift the truth from the error and extravagance so characteristic of this field. It will be our purpose in the pages which follow to present and evaluate, so far as may be possible, the more important experimental investigations bearing on the significance of anatomical and other alleged signs of aptitude.

VALIDITY OF CHARACTER JUDGMENTS BASED ON PHOTOGRAPHS

Probably the most widespread of all beliefs is that the character of a person is revealed in his face. That very definite impressions of character are produced in the observer by certain physical characteristics of the face of the one observed, no one can deny. Just what the facial characteristics are which produce these impressions, the observer is usually at very much of a loss to state. The degree of truth (if any) in this belief is a matter of much practical importance. If these impressions are really significant, they ought to be of enormous value in forecasting the aptitude and other behavior of persons with whom we come in contact every day.

This problem may be attacked experimentally from a number of different points of view. The first one which we shall consider is suggested by the extensive practice of requiring applicants for positions who do not appear in person to accompany their written application with a portrait photograph. While the photograph has certain other values, such as identification, the practice is certainly due in part to the persistent belief that even a representation of the face gives valuable information as to the kind of behavior which may be expected from the person himself. The first investigation of the validity of character judgments based on photographs was performed by Lucy G. Cogan and reported

by H. L. Hollingworth (33). A second study, performed under the present writer's direction by Florence E. McCabe (52), duplicated the technique of the earlier study so far as the experimental data are concerned. The only difference was that Miss McCabe had a larger number of subjects and had access to more adequate methods of statistical treatment for her experimental results than were in use when the investigation of Miss Cogan was performed. We shall accordingly describe the investigation as conducted by Miss McCabe, introducing the results of Miss Cogan for purposes of comparison.

Miss McCabe was fortunate in completing her study with 40 subjects, an unusually large number for an investigation of this kind. The subjects were university women, all members of the same sorority. These individuals were chosen as subjects of the investigation because in a sorority each person knows all the others with an intimacy hardly ever equaled in any other group of similar size. Of the 40 subjects, 20 who were regarded as the most reliable and as having the closest acquaintance with the others were selected to rank the entire group of 40 on 10 different traits. It was assumed that the joint judgment of these 20 intimate acquaintances would come as near to the true character of the persons ranked as is ever likely to be attained.

As already stated, the technique of securing quantitative judgments on the various character traits of the 40 individual women was that known as "ranking." One of the 20 "judges" would be given 39 small cardboards cut uniformly to a size approximating that of a man's calling card, each card containing the name of one of the remaining girls in the experimental group. The card of the young woman doing the ranking was never included because it was feared she would judge herself on a somewhat different basis from the others. The judge was then required to sort the cards over,

choose the card of the person possessing the greatest amount of the trait being rated, — e.g., “Neatness,” — and place that card first on the list. Then the card of the person manifesting the next most neatness would be placed second in the series, and so on down to the most slovenly person in the group. In this way each of the 20 judges ranked the remaining women of the group on each of 10 traits, yielding 200 series of character judgments.

Meanwhile photographs had been taken of all 40 women by the same photographer, all in exactly the same pose and in the same costume — a loose smock. In posing for the photograph the subject had her head turned toward the camera just enough to give a three-quarter view of the face. The head, neck, and shoulders were shown. Glossy prints giving much detail were made without retouching the negatives. The photographs were $2\frac{3}{4}$ inches by $3\frac{3}{4}$ inches in size. They were mounted on dull black cards which extended a scant quarter inch beyond the print.

These photographs were ranked on the same 10 traits as the originals of the portraits had already been, by a second squad of 20 judges entirely unacquainted with the persons whose photographs were being ranked. This second group of 20 judges was also made up of university women. Through their coöperation there were thus secured 200 additional series of ranks, paralleling those secured from acquaintances. The entire 400 series of ranks thus obtained were transmuted into linear units on a 10-point scale by means of the table in Appendix I, before being subjected to statistical analysis.¹

That the average judgment of a group of judges will agree fairly well with that of the average judgment of a second group of judges has frequently been observed in studies of this kind. If the judgments were mere random guesses,

¹ For an explanation of the necessity for this procedure, see pages 383 ff.

the two groups would correlate zero with each other. In the present investigation the pooled judgments of the first 10 acquaintance judges were correlated with the combined judgments of the second 10 acquaintance judges for each character trait. Similar correlations were computed for the first 10 and the second 10 stranger judges who made the character rankings from the photographs alone. The correlations are shown in parallel columns in the first part of Table 10. These figures show that the pooled judgments of the character traits made by close acquaintances are distinctly reliable. They also show very clearly that whether

TABLE 10

SHOWING THE CORRELATIONS BETWEEN POOLED JUDGMENTS OF THE FIRST AND SECOND HALVES OF THE JUDGES FOR VARIOUS TRAITS (After McCabe '26)

TRAITS	CORRELATION BETWEEN POOLED JUDGMENTS OF FIRST TEN AND SECOND TEN JUDGES		CORRELATION OF POOLED JUDGMENTS OF TWENTY JUDGES WITH TRUE MEASURE OF THE VARIOUS TRAITS ¹	
	When Judged by Associates on Basis of Acquaintance	When Judged by Strangers on Basis of Photographs	When Judged by Associates on Basis of Acquaintance	When Judged by Strangers on Basis of Photographs
Neatness . . .	+.93	+.90	+.98	+.97
Conceit . . .	+.88	+.76	+.96	+.93
Sociability . .	+.89	+.82	+.97	+.95
Humor . . .	+.81	+.84	+.94	+.96
Likability . .	+.83	+.82	+.95	+.95
Intelligence . .	+.92	+.66	+.98	+.89
Refinement . .	+.87	+.80	+.96	+.94
Beauty . . .	+.96	+.91	+.99	+.97
Snobbishness .	+.90	+.81	+.97	+.94
Vulgarity . .	+.80	+.63	+.94	+.88

¹ These values are derived from those of the first two columns by the formula $\sqrt{\frac{2r}{1+r}}$. A true measure is defined as the pooled or averaged judgments of an infinite number of independent judges as good as those actually employed.

there is any validity to character judgments based merely on photographs, there certainly is considerable unanimity in such impressions. This tendency to consistency of judgment comes out still more clearly in the second two columns. These represent the correlation that the pooled judgments of all 20 acquaintance judges and all 20 stranger judges would show if correlated with a true measure of whatever is being measured — i.e., with an average from an infinite number of independent judgments as good as theirs.

But the crucial question remains: Do the character judgments of strangers based on photographs reveal the true characteristics of the subjects as indicated by the joint and highly consistent opinions of those who know them best? This is shown by the correlations in Table 11. The first column shows the results of Miss McCabe's investigation here reported, and for comparison in a parallel column are given the results of Miss Cogan's investigation as reported by Hollingworth.

Hollingworth had concluded from Cogan's results that

TABLE 11

SHOWING THE CORRELATIONS BETWEEN REAL CHARACTER TRAITS AND JUDGMENTS OF CHARACTER BASED ON PHOTOGRAPHS

TRAITS	McCABE'S RESULTS	COGAN'S RESULTS (From Hollingworth)
Neatness	+.07	+.05
Conceit	+.21	+.19
Sociability	+.12	+.29
Humor	-.07	+.33
Likability	+.15	+.38
Intelligence . . .	+.40	+.51
Refinement . . .	+.17	+.51
Beauty	+.61	+.55
Snobbishness . . .	+.32	+.56
Vulgarity	+.10	+.65

while some of the correlations were fairly high, they were useless as a practical device because it required such a large number of judges before accurate enough judgments could be secured from the photographs to be of any value. Despite Hollingworth's pessimism the present writer had been much impressed by the size of the coefficients in the lower half of Cogan's column, since they compare in this respect very favorably indeed with the best psychological tests. It was accordingly something of a disappointment when Miss McCabe's coefficients (shown in the first column) were computed. It is true that all but one of these latter are positive. Only one, however, is high. This one is beauty. The high correlation with beauty is to be expected, since the thing itself, except for coloring, is recorded by the camera. Only one other coefficient (intelligence, + .40) is of any size. The results as a whole certainly look very bad for the judgment of character on the basis of photographs.

RELATION OF ACADEMIC APTITUDE TO CHARACTER JUDGMENTS
BASED ON PHOTOGRAPHS

There still remains, however, the one fairly high correlation — that with intelligence. One important form of intelligence is manifested in academic aptitude and is measured with fair approximation by the average of academic marks. The grades of these 40 subjects were accordingly obtained and averaged. Against these scholastic averages were correlated both the ratings by acquaintances on the true character traits and the ratings by strangers on the basis of the photographs. It was hoped that in the case of the photographs not only intelligence but some of the other traits might also be related sufficiently closely to scholastic aptitude to make it possible to combine several of them in such a way as to make a valuable supplement to present collegiate academic-aptitude tests. The results are shown in Table 12.

TABLE 12

SHOWING THE CORRELATION BETWEEN THE VARIOUS CHARACTER JUDGMENTS
AND ACADEMIC APTITUDE AS SHOWN BY UNIVERSITY MARKS

TRAITS	CORRELATION OF ACADEMIC MARKS WITH TRAITS AS JUDGED BY ASSOCIATES ON BASIS OF ACQUAINTANCE	CORRELATION OF ACADEMIC MARKS WITH TRAITS AS JUDGED BY STRANGERS ON BASIS OF PHOTOGRAPHS
Neatness	+ .37	-.15
Conceit	+ .07	-.11
Sociability	+ .04	+ .03
Humor	+ .16	+ .22
Likability	+ .21	+ .05
Intelligence	+ .74	+ .11
Refinement	+ .41	-.02
Beauty	-.15	-.16
Snobbishness	-.03	-.24
Vulgarity	-.32	-.05

The most noteworthy feature of the correlations with the traits as judged by associates is the very high correlation of .74 between marks and estimated intelligence. It need not be inferred from this, of course, that these women as judges of intelligence had any great insight into the constituents of academic success. This correlation is much more likely to reflect the fact that the girls tended to be rated on intelligence largely according to the kind of grades they were known to receive.¹

It must be admitted that an inspection of the second column of Table 12 gives scant grounds for the hope that anything of value, at least with respect to academic aptitude,

¹ It may be observed incidentally that beauty, both as judged by associates or by strangers on basis of photographs, shows a small negative correlation with marks. This tends to refute the belief often expressed that handsome girls tend to receive better marks than plainer ones for the same quality of work. This is supported further by the fact that intelligence correlated with beauty -.05, when both traits were judged by acquaintances. On the other hand, strangers judging photographs showed a correlation between estimated intelligence and beauty of + .41.

can be secured from photographs. The correlation with intelligence has disappeared when the photographic judgments are checked against an objective criterion. The most probable explanation of the correlation of .40 is that certain casts of features have somehow come to be regarded as indications of intelligence, that the judges of the photographs were influenced by this mainly, but that the associates combined this to a certain extent with their knowledge of reputation for marks and other activities.

HAVE PHYSIOGNOMISTS SPECIAL SKILL IN JUDGING CHARACTER
FROM PHOTOGRAPHS?

It might be true, of course, that even though the average of 20 intelligent university women had not been able to isolate anything from the McCabe photographs, possibly an expert specially trained or gifted might still be able to accomplish something of significance in this direction. With this in mind the author approached a phrenologist of national reputation, one who makes a practice of analyzing character from newspaper portraits, with the proposal that he rate the same photographs used above, with the understanding that in case it turned out badly his identity would not be revealed. It was found impossible to secure his coöperation. This refusal was particularly unfortunate, since here was an excellent opportunity to demonstrate his art if he really possessed any. On the other hand, if he had failed it would have been equally decisive, because no one has a greater reputation in this respect than this particular man. However, his failure to coöperate occasioned no surprise. This seems to have been the universal experience of psychologists in attempting to test out the skill of individuals claiming special powers of analyzing character from external signs. But in view of the fact that these claims have never been substantiated under scientifically controlled conditions, the

refusal to coöperate when such an opportunity is presented is itself highly significant.

VALIDITY OF JUDGMENTS OF PRACTICAL INTELLIGENCE
BASED ON PHOTOGRAPHS

The second study which we shall consider concerns the validity of judgments of practical intelligence based on photographs. It is reported by L. Dewey Anderson (1). The portrait photographs of 69 of the more important executive employees of a large department store were bound in a company annual. Twelve graduate students went through this book independently and sorted out what they regarded as :

1. The seven most intelligent persons.
2. The seven least intelligent persons.
3. The fourteen persons who are superior but not so good as the best seven.
4. The fourteen persons who are inferior but not so poor as the worst seven.

Appropriate scores were assigned to each of these groups (leaving out the middle 27), and an average of the scores assigned by all 12 judges was made. This was correlated with the scores made by the same subjects on an adaptation of the Army Alpha test. The two were found to correlate $+ .27$. Owing to the leaving out of the middle group of subjects in the compilation, this coefficient is certainly much too large, though there evidently was a trace of tendency to agreement.

These average photographic ratings on 31 of the store managers were next correlated with the average ratings by each other as to their general value to the firm. Here again the correlation, though small, was found to be positive, $+ .22$. This coefficient also must be discounted considerably for the same reason as the first correlation mentioned.

EVIDENCE OF BLOND AND BRUNETTE COLORING AS SIGNS
OF TEMPERAMENT

Since blondness and brunetteness are so easily recognized by an employer in an applicant for a position, the claims of the striking relation between coloring and character traits have attracted considerable attention. Two very similar studies have been carried out to test this theory. In one of these investigations, conducted by Paterson and Ludgate (63), 94 students in a class in psychology were asked to select four persons of their acquaintance, two of whom were pronounced blonds and two decided brunettes. They were then asked to record on a suitable form, by marking with plus and minus signs, whether each of these persons possessed each of a long list of traits claimed to be characteristic either of blonds or brunettes. The percentages of the 187 blonds and 187 brunettes said to possess each of the traits are shown in Table 13. If the claims were true, we should expect to find percentages approaching 100 in the first half of the first column and the lower half of the second column, with zero or at least very small percentages in the remaining half columns. As a matter of fact, we find the differences between the corresponding entries in the two columns very slight. An examination of the two columns for any tendency to agreement reveals the fact that out of the 26 pairs of percentages, 20 (or 77 per cent) are slightly in favor of the claims. Chance alone would have produced about 50 per cent.

A second study of a very similar nature was carried out by Kenagy. He asked 38 sales executives each to rate the four best-producing salesmen well known to them on the same 26 traits shown in Table 13. It turned out that of the 152 best-producing salesmen, 82 were brunettes while only 70 were blonds, which goes decidedly against the current claims. The results of this investigation as summarized

in the last column of Table 13 shows that 18 out of the 26 traits (or 69 per cent) tend slightly to favor the claims. This is in substantial agreement with Paterson and Ludgate. This may mean that there is a mere trace of a tendency for the claims to be true. If this is the case the tendency is much too feeble to be of any practical value, at least unsupported by other evidence. The probability is, however, that scattered through the raters in both investigations were a few individuals familiar with the Blackford claims and that these individuals were either consciously or unconsciously influenced by this familiarity when making their ratings.

VALIDITY OF CHARACTER JUDGMENTS BASED ON SEEING
THE SUBJECTS IN PERSON

But there remains still another possibility concerning physiognomy which is worthy of consideration. There is the possibility that when we are actually judging the character of a person, features other than those recorded by the camera may be significant. The most obvious factor in this connection is what is commonly spoken of as "facial expression." By this is here meant the behavior or movements of the features under the varying but more or less casual stimulations of social contact. The photograph usually shows the face in repose. Even if the features happen to be caught in any special expression, no sequence of expression can be shown; much less is there any revelation of what evoked the expression. In this connection an investigation carried out by Cleeton and Knight (15) is of interest.

In connection with an investigation of the claims of phrenology, they examined the possibilities of judging character when the judge views the subject face to face. They had as subjects 28 university students chosen in about equal numbers from two national sororities and a national fraternity. The members of each of these three groups were

TABLE 13

SHOWING THE PERCENTAGES OF BLONDS AND BRUNETTES POSSESSED OF TRAITS CLAIMED AS CHARACTERISTIC OF BLONDS AND BRUNETTES, RESPECTIVELY

ALLEGED BLOND TRAITS	RESULTS OF PATERSON AND LUDGATE			KENAGY'S RESULTS SUMMARIZED AS TO AGREEMENT WITH CLAIMS
	Per Cent of 187 Blonds Possessing Trait	Per Cent of 187 Brunettes Possessing Trait	Agreement (+) Disagreement (-) with Claims	
Positive . . .	81	84	—	+
Dynamic . . .	63	64	—	+
Driving . . .	49	50	—	—
Aggressive . . .	62	56	+	—
Domineering . . .	36	36	?	+
Impatient . . .	56	51	+	—
Active . . .	88	82	+	+
Quick . . .	70	68	+	+
Hopeful . . .	85	85	?	+
Speculative . . .	53	51	+	+
Changeable . . .	53	43	+	+
Variety loving .	66	62	+	+
ALLEGED BRUNETTE TRAITS				
Negative . . .	16	17	+	+
Static . . .	28	31	+	?
Conservative . . .	51	61	+	+
Imitative . . .	39	40	+	+
Submissive . . .	25	26	+	+
Cautious . . .	54	60	+	+
Painstaking . . .	56	61	+	—
Patient . . .	43	52	+	—
Plodding . . .	27	31	+	+
Slow . . .	20	24	+	+
Deliberate . . .	47	57	+	+
Serious . . .	58	72	+	?
Thoughtful . . .	67	70	+	+
Specializing . . .	52	45	—	—

then rated by 20 of their intimate associates in the respective organizations on the eight psychological traits shown in Table 14. Each of these groups was then placed on a stage in full view of 70 judges, each of whom scrutinized the subjects as much as he wished, and then rated them on the same traits as those on which they had been rated by their associates. The judges were strangers, but were persons accustomed to employing people, and included business men, school principals, employment managers, and others.

For the three groups, the average correlations between the character judgments of 20 intimate acquaintances on the one hand and of 70 observing strangers on the other are shown in Table 14. While the coefficients here shown are all very small, it can hardly be accidental that they are all

TABLE 14

CORRELATION EXISTING BETWEEN RATINGS OF 20 CLOSE ASSOCIATES AND RATINGS OF 70 CASUAL OBSERVERS ¹

Judgment	+.32
Intelligence	+.02
Frankness	+.21
Will power	+.26
Ability to make friends	+.18
Leadership	+.31
Originality	+.32
Impulsiveness	+.20

positive. There is no particular indication, however, that any special advantage is gained from seeing the subjects in three dimensions over seeing the photographs, since the correlations are no higher than those obtained by Cogan and by McCabe on rather similar traits.

¹ After Cleeton and Knight (15).

EVIDENCE AS TO THE SIGNIFICANCE OF CONVEX AND
CONCAVE PROFILE

A somewhat elaborate investigation as to the significance claimed for the profile (page 113) was carried out by Alice L. Evans¹ (19). Miss Evans employed as subjects 25 university women who were members of the same sorority. The members of a sorority were chosen as subjects of the investigation, as in McCabe's study, because of the intimacy of acquaintance characteristic of such groups. All of the individuals who had not been members long enough to be well known to all the others were excluded. Each girl ranked the remaining 24 on each of the seven traits shown in Table 15, using cards as described above (page 115). Thus each girl was alternately subject and judge. The rankings for the several traits made by any given judge were separated by a week or more to prevent the rankers, so far as possible, from confusing the various traits while supposedly ranking a particular one. When all subjects had ranked the others on any particular trait, these rankings were pooled to get a single joint judgment as to the amount of the trait exhibited by each individual. This was considered the best procurable measure of the trait in question.

As pointed out above in connection with McCabe's study of photographs, the individuals of a group of this kind show considerable agreement in their ratings of the various individuals with respect to the several traits. This is shown by the fact that there is very striking agreement between the rank of the 25 subjects on a given trait yielded by the pooled ranks from the first 12 of the subject-judges when compared with the pooled ranks from the remaining 13 subject-judges. The pooled rankings thus obtained from the two groups of

¹ Most of the correlations have been recalculated by more exact methods than were in use at the time that Miss Evans did her work.

subject-judges were correlated, yielding the results shown in the following table:

TABLE 15

SHOWING THE CORRELATION BETWEEN THE POOLED JUDGMENTS FROM THE TWO HALVES OF THE SUBJECT-JUDGES FOR THE VARIOUS TRAITS

Optimism	+.75
Activity	+.80
Ambition	+.90
Will power	+.82
Domination	+.87
Popularity	+.44
Blondness	+.93

It is noteworthy that in this table the coefficients for the elusive character traits, except that for individual popularity, compare rather favorably with those for a highly objective trait such as blondness.

In order to determine the convexity of the various faces, a special instrument was devised by the writer. It is shown in Figure 22. The instrument was so constructed that it could be applied to the face of a person and the angle of convexity could be read off directly in degrees of arc. The instrument was calibrated so as to register zero when placed on a plane surface like a table top. In order that no possibilities should be overlooked, the convexity of the profiles was measured in a number of independent ways: whole face, upper face, lower face, both with and without the nose. The measures spoken of as "without nose" were taken with the divided middle foot of the instrument resting on the face astride the nose, at the base of the latter. The height of the forehead was measured with a tape. Lastly, one physiognomic trait (blondness) was obtained by having the subjects rate each other exactly as was done in the case of the character traits.

With these various measures it is possible to test the accuracy of the alleged relation between blondness and profile

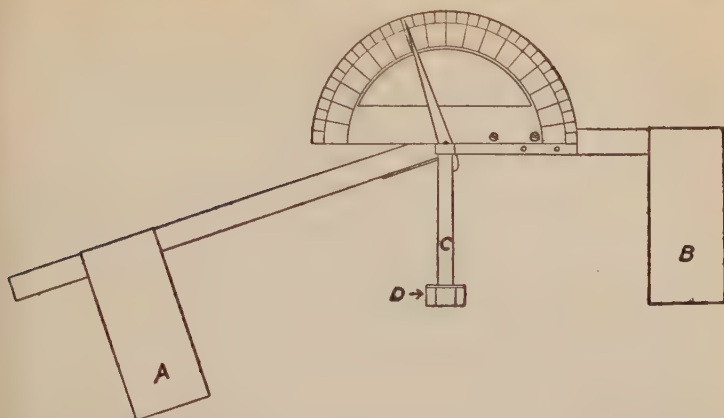


FIG. 22. Showing device for measuring convexity of the profile. It is set for a convexity of 10 degrees. The legs *A* and *B* at the ends of the arms may be slid along the arm to adjust to any length of face. The leg *C* is double so as to stand astride the nose when desired. The detachable bridge *D* connects the two legs at *C* in such a way as to rest on the nose if desired.

convexity on the one hand, and the various character traits on the other. The various final rankings of Miss Evans's data were converted into linear scores by means of the table in Appendix I, after which the critical correlation coefficients were computed by the Pearson formula. They are shown in Table 16. If the claims suggested by the quotations given above (page 113) were sound, we should expect a high positive correlation between the physiognomic traits and the first five of the character traits. An inspection of the table, however, is somewhat disappointing. In the first place, blondness correlates neither with convexity nor with the various character traits. For the most part, convexity also seems unrelated to the various character traits. A possible exception to this seems to be convexity of the lower face (chin to eyebrow without nose), which yields a curiously consistent series of relatively large positive correlations with

TABLE 16

SHOWING THE CORRELATION BETWEEN VARIOUS PHYSIOGNOMIC TRAITS
AND A NUMBER OF CHARACTER TRAITS (Based on Evans's data)

PHYSIOGNOMIC TRAITS	CHARACTER TRAITS						BLOND- NESS
	Optimism	Activity	Ambition	Will Power	Domination	Popularity	
Convexity, whole face with nose	+ .10	-.05	-.17	-.13	-.11	-.03	-.20
Convexity, chin to eye- brow, with nose . .	+ .13	+ .01	-.13	+ .13	-.08	-.11	+ .03
Convexity, whole face without nose . . .	+ .02	-.24	-.17	-.11	-.13	-.27	-.04
Convexity, chin to eye- brow, without nose	+ .37	+ .39	+ .33	+ .34	+ .24	+ .17	+ .03
Convexity of upper face, with nose . .	-.06	-.08	+ .04	+ .06	+ .08	-.17	-.02
Height of forehead from eyebrow to hairline	-.17	-.29	-.23	-.39	-.22	-.10	-.21
Blondness	-.26	-.02	+ .05	+ .28	+ .14	+ .03	

all of the character traits. Because of the small number of subjects the probable error of these coefficients is rather large, but the correlations are nevertheless suggestive. A second possible exception lies in the series of negative correlations between the various traits and the height of the forehead from eyebrow to hair line. If these coefficients may be taken at face value, they indicate that *low* foreheads tend to indicate optimism, activity, will, etc.; but here again we must reserve judgment because of the small number of subjects. These coefficients are sufficiently suggestive to warrant further investigation.

DIMENSIONS OF THE HEAD AS SIGNS OF APTITUDE

It is very general indeed in the dogmatic literature on the subject for physiognomic and phrenological claims to be associated. An experimental investigation which similarly combined the two was carried out in the Wisconsin laboratory by Elsie B. Sherman (76). The subjects of this investigation were 78 young men, all freshmen of the Engineering College. Numerous measurements of the skulls and faces of these men were made by a specially designed radiometer.

This instrument (Fig. 23) resembles in some respects earlier devices of the same general nature. It is provided with suitably shaped fiber knobs (*A*, *A'*) to be inserted into the external meatus of each ear. The two ear pieces are supported by a semicircular rod (*B*), in the middle of which is a device for holding a piece of wood (*C*) upon which the subject bites. The mouth piece and the two ear pieces support the moving parts of the instrument rather firmly when measurements are to be taken. Mounted on the supporting framework just described and pivoted at the aural axis is the semicircular brass piece (*D*). This may be revolved freely about its axis and so around the head, the particular position that it occupies being registered in degrees on the angular scale at *E*. The scale and pointer are so placed with respect to each other that when *D* is exactly opposite the mouth board they register zero degrees. Mounted on *D* is a device (*F*) which contains a slot through which a special ruler (*G*) is inserted. This ruler is so calibrated that when pushed through the slot until its tip comes in contact with any point of a person's head, there may be read off directly the number of millimeters the point is distant from the center of the head at the aural axis.

The device *F*, while designed to be moved about on its semicircular support at will so that a great variety of dimen-

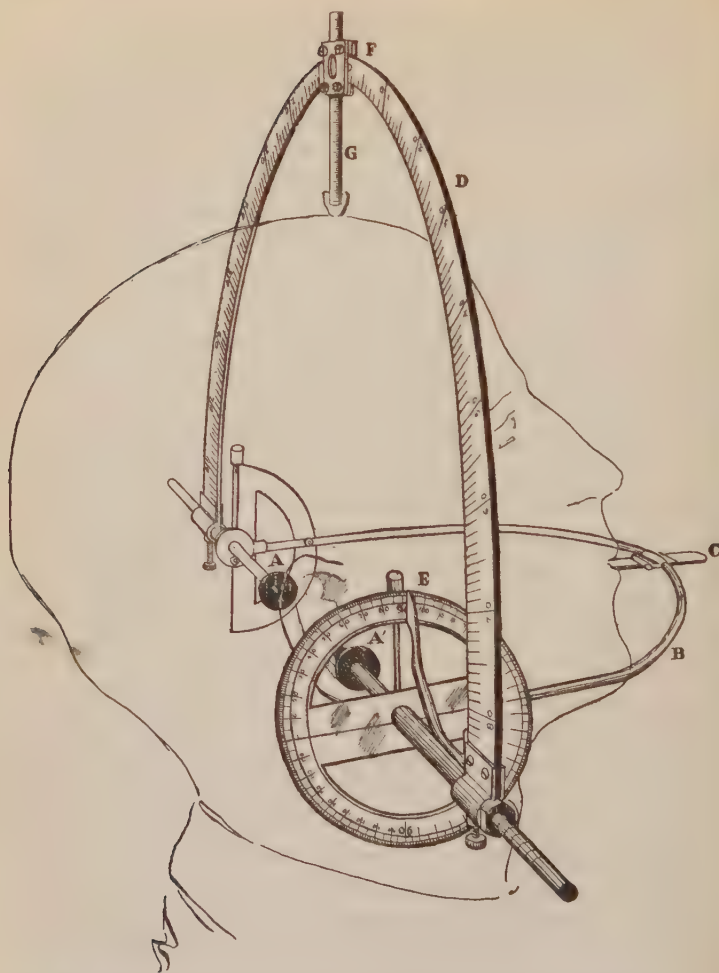


FIG. 23. Showing the instrument for making precise head measurements.
(Hull.)

sions might be measured, was throughout the present investigation maintained in the central position. The measures taken were consequently all in the median plane. They are listed in the first column of Table 17. The significance of various symbols which appear there will, for the most part, be apparent upon an inspection of Figure 24. It need only be explained that OO is the width of the head from meatus to meatus, that $SO = LO - FO$ and $RO = LO \div FO$. It should be added that the distance GH was not measured directly but was computed trigonometrically from other available measures, as was also the angle ADF . All of the linear values are in millimeters.

No attempt was made to compare the above measures with the various character traits alleged to be revealed by the dimensions of the head. It was felt that even though head measures might possibly be related to the character traits, they would have no practical value unless they were also significantly related to specific aptitudes. The aptitude chosen for investigation was that of mastering the curriculum of an engineering college during the first semester of the freshman year. The subjects in this curriculum were: Chemistry, English, Mathematics, Mechanical Drawing, Forge Shop Work, and Machine Shop Work. The criteria of aptitude were the final grades in the respective courses at the end of the semester. Intercorrelations among the six subjects of instruction revealed the fact that English, Mathematics, and Chemistry tended rather strongly to resemble each other and thus to form a psychological group. In a similar manner mechanical drawing and the two shop courses tended to form a second psychological group, presumably dependent more on abilities to do things with the hands. Accordingly, the first three studies were combined for one criterion and the second three for a second criterion. Lastly, all six studies were combined for a third criterion.

TABLE 17

SHOWING THE CORRELATION BETWEEN VARIOUS MEASURES OF THE HEADS OF 78 FRESHMEN ENGINEERS AND GRADES RECEIVED DURING THE FIRST SEMESTER (Computed from Sherman's data)

HEAD MEASUREMENTS	SCHOLASTIC-APTITUDE CRITERIA		
	Combined Chemistry, English, and Mathematics	Combined Drawing and Shop Courses	All Six Courses Combined
Angle <i>AOD</i>	+ .18	+ .17	+ .23
" <i>FOD</i>	- .14	+ .03	- .04
" <i>ADF</i>	+ .04	+ .03	+ .03
" <i>GOH</i>	- .27	- .16	- .26
Distance <i>OO</i>	+ .03	+ .07	+ .07
" <i>AO</i>	+ .08	+ .01	+ .04
" <i>BO</i>	- .04	- .08	- .09
" <i>DO</i>	+ .03	- .19	- .09
" <i>EO</i>	+ .01	- .10	- .05
" <i>FO</i>	+ .09	- .08	+ .01
" <i>GO</i>	+ .08	- .06	+ .02
" <i>HO</i>	+ .27	+ .26	+ .32
" <i>LO</i>	+ .25	+ .16	+ .23
" <i>MO</i>	+ .24	+ .23	+ .27
" <i>NO</i>	+ .28	+ .28	+ .34
" <i>PO</i>	+ .27	+ .18	+ .27
" <i>SO</i>	+ .19	+ .20	+ .22
" <i>RO</i>	+ .07	+ .14	+ .10
" <i>GH</i>	- .18	- .08	- .15

The correlations between each of the nineteen head measures and each of the three criteria are shown in Table 17. As might be anticipated, the most of the coefficients are small. We may note first the correlations involving two of the head measures which are of special interest because closely related to face measures yielding suggestive correlations for Evans. The first of these measures is the angle *ADF*, which corresponds roughly to Evans's angle of convexity for the lower face without nose (page 130). This measure yielded for Evans positive correlations ranging around + .30 for a

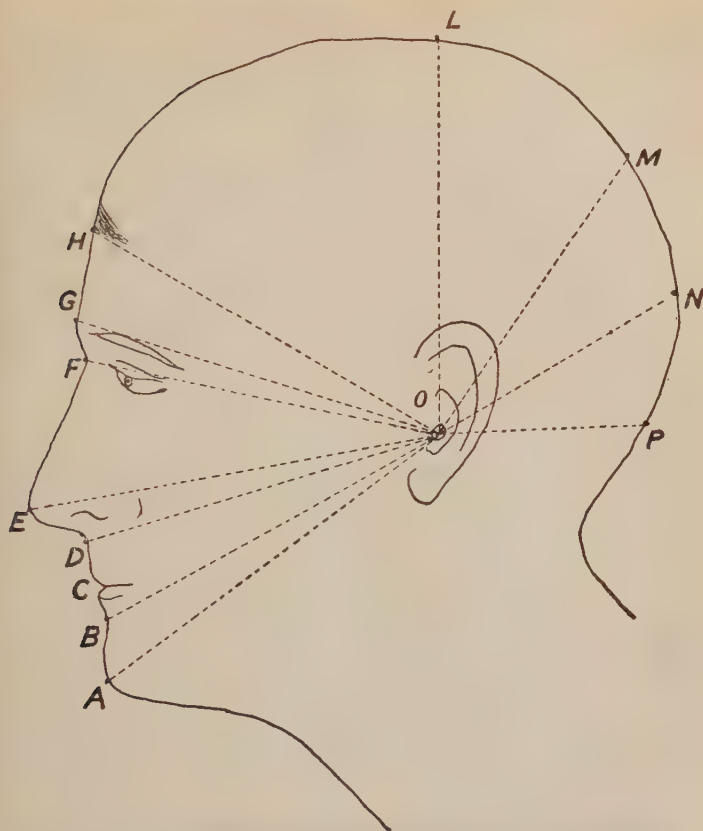


FIG. 24. Showing head measurements performed by Miss Sherman.

variety of dynamic character traits. These traits are such as might reasonably be expected to manifest themselves in academic achievement. Table 17 shows that while angle ADF yields positive correlations with academic performance, the coefficients are so small as to be negligible.

The second measure of special interest because of Evans's

work is the distance GH . This corresponds directly to the distance from eyebrow to hairline as measured by Evans, who found fairly large and consistent negative correlations with the dynamic character traits measured above. Sherman's results in this case tend to confirm those of Evans, since they yield consistent negative and fairly large correlations with academic achievement. If the results of these studies should be confirmed by further investigation, it will be an interesting case of the overturning of a popular belief which is to the effect that a high forehead indicates excellent academic aptitude. In short, it would necessitate a reversal of the current use of the expressions "highbrow" and "lowbrow."

By comparing Table 17 with Figure 24 it will be seen that, without a single exception, the measures from the ear to the surface of the skull over the brain yield positive correlations. There is excellent reason to believe that this is not accidental. An inspection of the correlations among these particular head measures showed that in many cases they were not so highly correlated among themselves as might have been expected. This suggested the possibility of combining them so as to make a joint correlation with the third criterion. It was found that when measures HO , LO , MO , NO , and PO were weighted in the best possible manner, they gave a joint correlation yield of .44.

Turning to the angles, it is seen that two of these have distinct correlations with our third criterion, angle AOD (on the lower face) yielding $+.23$ and angle GOH , $-.26$. These angles were found to be highly independent of the dimensions over the brain, especially GOH . This suggested the possibility of combining these two angular measures with the three strongest measures over the brain so as to form a kind of battery of anatomical signs. It was found that when combined in the best possible manner these five head

measures gave a correlation yield of .50 with academic marks. This is distinctly promising. Indeed, entire test batteries designed for predicting academic success of university students frequently do no better. While it can hardly be expected that head measures will consistently yield correlations of .50, these results suggest the desirability of further investigation. The formula by means of which the various head measures may be converted into a forecast of school marks is: $\bar{X}_0 = .01 X_1 - .04 X_2 + .28 X_3 - .11 X_4 + .26 X_5 + 76.9$.¹ In this formula, $\bar{X}_0$ is the most probable average mark as estimated by the formula; X_1 is the angle AOD ; X_2 is angle GOH ; X_3 is HO ; X_4 is MO ; and X_5 is NO . The last three are all measured in millimeters.

It cannot be denied that among academic psychologists there is a rather strong prejudice against placing any faith in the significance of anatomical signs as indicative of intellectual aptitude. This is a perfectly natural reaction to the extravagance, dogmatism, and even charlatanism of the phrenologists and physiognomists. It has been known for a long time, however, that there is a positive relation between various dimensions of the brain case and academic aptitude. The most extensive results in this field are those of Karl Pearson. He measured the skulls of 1011 graduates of Cambridge University and found a correlation of +.111 between length of head and a rough measure of academic success. Pearson also found a correlation of +.097 between head width and the same rough measure of academic success. Sommerville (78), working with 117 students scattered through various years at Columbia University, found a correlation of +.09 between head length and a composite measure of academic success. With the same criterion she secured a correlation with academic success of +.138 in the

¹ For an explanation of the use of forecasting formulæ, see pages 179 ff. For the methods of deriving such formulæ, see Chapter XIV.

case of head height and one of $+ .07$ in the case of head width. There is thus no longer any question but that there is a positive though small correlation between certain dimensions of the skull and the academic aptitude of university men.

The usual objection to the use of such measures for purposes of practical prognosis has been voiced by Pearson. He holds quite correctly that no one of these measures has strong enough forecasting power to be of any practical value when standing alone. This, however, is not a fair test of value. By the same method practically all forms of single psychological tests would also be found valueless. Sherman's results suggest that when accurate head measures are combined as we are in the habit of combining psychological tests, they may compare rather favorably with the latter. It will be recalled that by combining five of Sherman's head measures, a multiple-correlation coefficient of $.50$ was obtained. No doubt this correlation should be reduced to a certain extent on account of the well-known tendency for such coefficients to exaggerate the correlation to be obtained from combining in this way a number of measures to make aptitude predictions. Nevertheless, there remains a distinct presumption that certain dimensions of the head, when combined with other indicators of aptitude such as samples of test behavior, may contribute unique and substantial increments to the prognosis of aptitudes.

THE ENDOCRINE GLANDS AND SOMATICO-BEHAVIOR TYPES

A distinctly modern phase of physiognomy and anatomical signs in general must now be considered. This new movement resembles the older one in being largely qualitative in its methods and in depending upon extreme cases, often of the pathological sort, to support its claims. The claims are often rather extravagant and ill supported by evidence.

It represents, however, an enormous advance over the traditional physiognomy in many ways. The majority of its exponents are persons of more or less recognized standing in science, they take a distinctly scientific point of view, and some of them actually check up their conjectures by systematic and even quantitative investigation. Lastly, they have a theory which, whether true or false, is at least intelligible.

It has long been known that there are certain glands of the body which are not provided with ducts through which to discharge their secretions (as is the case with the salivary glands, the liver, etc.) but, instead, throw their secretions directly into the blood stream. These are the so-called ductless or endocrine glands. Some of the more important of these glands are the thyroid, the pituitary, the adrenals, and the gonads. The interactions among these glands, and their actions upon the growth and functions of the body, are important but exceedingly complex. Despite a vast amount of research devoted to the subject, comparatively little is yet known about their influence upon other parts of the body.

The interest of the aptitude psychologist in the endocrine glands has been stimulated by indications that an overactivity or underactivity of a certain gland may influence the rate and character of the growth of certain parts of the body, while at the same time producing an equally characteristic influence upon the general behavior and temperament of the person. If such forces be assumed as operating on a large scale, it would be easy to understand how an anatomical trait such as a special shape of face or proportion of the legs to the torso might indicate a characteristic temperament or behavior tendency. It is now well known, for example, that cretinism, a special form of feeble-mindedness which also shows very characteristic accompanying forms of physical peculiarities, is the result of defective action of the thyroid gland. It has also long been known that removal of the

gonads by castration during the period of growth also profoundly influences both bodily and character development. Extending this line of reasoning, a number of writers have developed more or less elaborate systems of parallel physical and temperamental types. One of the more prominent of these writers is E. Kretschmer.

KRETSCHMER'S SOMATICO-BEHAVIOR TYPES

Kretschmer (48), like so many of the writers in this particular field, takes his point of departure definitely from the extreme forms of behavior deviation of psychopathology. He isolates three physical types:

1. Asthenic type, tall, slender, and generally feeble.
2. Athletic type, strong, vigorous shoulders with muscles and bones little obscured by fat.
3. Pyknic type, inclined to be fat, round, and plump.

His two main types of temperaments are:

1. Cycloid, partaking of the nature of manic-depressive madness, especially of the manic phase.
2. Schizoid, partaking of the nature of dementia præcox, the tendency to live within one's self, unsociable, timid, shy, eccentric.

As definite evidence indicating the significant relation between the physical types on the one hand and the behavior types on the other, Kretschmer gives a tabular distribution of his clinical cases. A part of his table, illustrating clearly his view, is reproduced as Table 18. If these results are not seriously influenced by Kretschmer's own theories in making the classifications, they would seem to indicate a very strong correlation between the pyknic physical type and the cycloid disease type, on the one hand, and the asthenic and athletic physical types and the schizoid mental disease type, on the other. How far the individuals represented in the table

TABLE 18

SHOWING THE RELATION BETWEEN PHYSICAL TYPES AND MENTAL DISEASE TYPES (After Kretschmer)

PHYSICAL TYPE	MENTAL DISEASE TYPE	
	Number of Cyloids	Number of Schizoids
Asthenic	4	81
Athletic	3	31
Pyknic	58	2

showed both the same physical types and the same behavior temperaments *before* the onset of the disease is not known. Unfortunately, from the point of view of forecasting, this detail is of central importance.

As integral parts of Kretschmer's general anatomical types he isolates parallel facial types. In particular, he distinguishes an asthenic facial type which, of course, is supposed to be associated with the schizoid or non-aggressive temperament. This facial type is claimed to have a relatively long nose and a short jaw producing a receding chin. Dr. Sommerville, in connection with certain skull measurements on Columbia students, secured an index of this trait by dividing the length of the nose by the distance from the point of the chin to the angle of the jaw bone. An extreme asthenic face, by this measure, should give a large index, whereas a strong chin with medium length of nose should produce a small index. He found this index of 117 subjects correlating — .04 with university marks and — .17 with extra-curricular successes. While low, these correlations, particularly the latter, tend to give a little color to the natural expectation that the association of extreme tendencies reported by Kretschmer may hold to a certain extent among normals. This inference is based on the assumption that the greater the inclination

toward the schizoid temperament the more retiring and un-aggressive in extra-curricular activities and hence the less successful. Owing to the nature of the facial index used, a negative coefficient should be expected.

A second study designed to test Kretschmer's claims is reported by Mohr and Gundlach (58), based on a prison population. Upon the whole they find a tendency to agreement.

NACCARATI'S MORPHOLOGICAL INDEX AND ACADEMIC APTITUDE

A second claim of somatico-behavior parallelism deriving its theoretical basis from endocrinology has been put forward by S. Naccarati (60). His view is, in brief, that persons who at about 20 to 25 years of age are relatively slender in build will have greater intellectual capacity than those more stocky in build, particularly in the trunk region. The context indicates that by "intelligence" is here meant academic aptitude. In its complete form this index is based upon a very elaborate series of measurements which are calculated to yield reliable measures of the length of limbs and the volume of the trunk respectively. The former value is then divided by the latter, which yields the true morphological index. As a rough but simple approximation to the morphological index, Naccarati in some of his investigations merely found the height and weight of the subjects, taking as his index the height in centimeters divided by the weight in kilograms. This he calls the height-weight ratio. His criterion of intelligence was the score made on the Thorndike Intelligence Examination.

In his first and most important report (60) Naccarati employed several groups of subjects. For purposes of computation certain of these groups were combined. Since Naccarati has published his original data, we have taken the trouble to recompute his coefficients and incidentally to

TABLE 19

SHOWING FOR NACCARATI'S VARIOUS GROUPS OF SUBJECTS THE CORRELATIONS AMONG THE SEVERAL VARIABLES (Computed from Naccarati's published data)

GROUP OF SUBJECTS	NUMBER OF SUBJECTS	CORRELATED VARIABLES	<i>r</i>
I	221	Intelligence and Height	+ .03
I	221	" " Weight	- .199
I	221	" " H. W. ratio	+ .231
I	221	Height and Weight	+ .498
I	221	Intelligence and Height - 1.7 Weight	+ .25
II	75	" " Height	+ .025
II	75	" " Weight	- .324
II	75	" " Morphologic Index	+ .382
II	75	Height and Weight	+ .388
II	75	Intelligence and Height - 4.4 weight	+ .363

perform certain other computations not reported by him. These results (Table 19) indicate that of the two, weight is much more highly correlated with intelligence test score than is height. Clearly, whatever strength the height-weight ratio may have in forecasting intelligence test score is due chiefly to the negative relation with weight. This suggested the possibility of combining height and weight by addition, the latter being weighted in the best possible manner. The most effective weight in one case was found to be - 1.7, and in the other - 4.4. The correlation yields in both cases approximate very closely to those obtained by Naccarati's ratio method. These results would seem to indicate that it might at least be worth while to consider the physical weight of students as a factor in estimating academic aptitude.

The interpretation of Naccarati's results is somewhat complicated by the fact that since the publication of his study two investigations following a somewhat similar procedure have failed to find any confirmation. Sommerville (78), to

whom reference has been made in connection with head measures (page 141), correlated Thorndike intelligence scores with an index produced by dividing the sum of arm and leg length by sitting height. It was assumed, apparently, that sitting height would correspond to volume of trunk somewhat as weight did in Naccarati's work. This is questionable. A correlation of zero was obtained, though a negative correlation should have been expected if the tacit assumption regarding sitting height had been sound. Sommerville also divided the weight times 1000 by the height squared. This quotient correlated $-.05$ with Thorndike intelligence. She then divided weight times 1,000,000 by the height cubed. This correlated $-.10$ with Thorndike intelligence. The signs of these correlations are in harmony with Naccarati's results, since Sommerville's ratios were inverted with respect to his, in so far as they were comparable. The coefficients are very small, however.

A second study which is much more closely comparable with those of Naccarati has been reported by Edna Heidbreder (29). She correlated the group test scores of 1000 students at the University of Minnesota with their height-weight ratio, apparently in complete conformity with Naccarati's simpler technique. She reports a correlation of almost exactly zero. Since this coefficient is based on such an impressively large number of subjects, it serves seriously to call in question the significance of the height-weight ratio as a sign of academic aptitude. But we have already seen above that Naccarati himself obtained much lower correlations with the height-weight ratio than with his morphological index.

It may still be that the more elaborate method will yield results when the simpler one fails. This is indicated by an investigation reported by Sheldon (75). Using Naccarati's morphologic index, he obtained a correlation of $+.11$ with

university grades and a correlation of $+ .14$ with a group scholastic-aptitude test. Since 434 subjects were used, these results are highly reliable and probably indicate about the extent of the relation between the morphologic index and scholastic aptitude.

DOES THE SHAPE OF THE HAND REVEAL CHARACTER?

A special group of alleged anatomical signs often associated in the literature of characterology with physiognomy and phrenology relates to the shape and proportions of the hand. This body of lore is called *chirognomy* or hand reading.

Chirognomy professes to discover the character of a person from the form and proportion of the hand. It thus differs from *chiromancy*, or palm reading, which professes to read both the past and the future of the subject in the lines of the palm. One writer on the subject of chirognomy claims to have given private lessons in the study of the hand to the heads of 275 business establishments in New York, 135 in Boston, and 342 in Chicago. "All these men," he states, "were large employers of labor and what they principally wanted was to have some help beyond their own judgment in dealing with those with whom they came in contact in the ordinary course of their business careers."

An experimental investigation of the claims of the chirognomists was carried out by Dorothea MacLaurin (55). The general plan of her study was to secure some reliable character ratings from 30 members of a sorority, then make careful measurements of the subjects' hands according to the principles laid down in works on chirognomy, and lastly to compute correlations to see whether the shape of the hand really was associated with the character traits as claimed. After consultation of numerous works on the subject, five traits were found upon which there was considerable agreement among the writers and in connection with which the alleged

sign in the hand would permit of objective measurement. It is claimed, for example, that the longer the first finger as compared to the second, the more ambitious a person is; that the farther a person can bend his fingers backward, the keener is his mind; that the longer the fingers in proportion to the length of the palm, the stronger is the tendency to impulsiveness, and so on.

The main results of this investigation are summarized in Table 20. Upon the whole, these coefficients do not indicate a very hopeful future for chiromnomy. Indeed, the extremely low correlation values and the high P. E.'s, coupled with the utter lack of any theoretical reasons for expecting any special relation between the shape of the hand and the character of the individual, pretty definitely remove chiromnomy from the field of practical aptitude prognosis.

TABLE 20

SHOWING THE CORRELATION BETWEEN CERTAIN MEASURED PROPORTIONS OF THE HAND AND THE CHARACTER TRAITS ALLEGED TO BE INDICATED BY EACH (Computations based on MacLaurin's data)

HAND TRAIT	CHARACTER TRAIT	CORRELATION COEFFICIENT (r)
Difference in length between first and second finger	Ambition	+ .19
Extent of backward flexion of the hand measured in degrees	Keeness of mind	+ .13
Degree of taper of fingers	Refined sensibility	+ .16
Length of fingers divided by length of palm	Impulsiveness	+ .29
Length of thumb divided by length of palm and second finger	Force of character	- .23

EVIDENCE AS TO THE REVELATION OF CHARACTER BY
HANDWRITING

Passing from the physical aspects of the hand, we come to the consideration of graphology, which concerns the behavior of the hand, as seen in handwriting. It is claimed that the character of an individual is revealed in great detail by his writing. Like physiognomy, graphology has a long history and has included among its adherents many eminent persons. It is only very recently, however, that any really critical scientific evidence bearing on the truth of its claims has become available.

Perhaps the first scientific attempts along this line were made by Alfred Binet (4) during the preliminary gropings which finally led to the construction of his now famous psychological tests. On one occasion he had graphologists sort out according to the sex of the writer 180 envelopes addressed to him in about equal numbers by men and women. In this task chance guesses would of course yield about 50 per cent correct decisions. All the graphologists run much ahead of chance, one as high as 79 per cent of correct judgments. In another study Binet obtained samples of writing from 37 individuals of recognized intellectual eminence, such as Renan and Bergson. With each of these he paired the writing of a person of similar education and general social level but of very mediocre intelligence. Several different graphologists were asked to state which of each pair of scripts was written by the more intelligent person. The scores of all were distinctly superior to chance, the range being from 92 per cent to 61 per cent, where chance alone would have yielded 50 per cent. Binet accordingly came to the conclusion that there was a certain amount of truth in the claims of graphologists. It is not without significance, however, that when he published his epoch-making tests, no trace of graphology was included.

Binet's investigations were designed to test the skill of particular graphologists. It is to be observed that skill may exist, although the rationalizations of the skill — i.e., the reasons given for various decisions — may be false. Two investigations have attempted to test in an objective manner the accuracy of the system by which graphologists claim to delineate character from handwriting. Hull and Montgomery (41), after comparing a large number of works on graphology, selected six character traits concerning which there was fair agreement and which also were held to be associated with traits of handwriting that were susceptible of objective measurement. Seventeen university students belonging to a medical fraternity acted as subjects. These men ranked each other on the various character traits. They were also asked to copy a certain piece of prose, all on the same kind of paper, at the same desk and with the same pen. The various relevant proportions of each sample of writing were then measured with the greatest care. Where necessary, microscopic measurements were resorted to. These various measurements were then correlated by the Spearman rank method with the joint judgment of the character traits of the men. The resulting coefficients are shown in Table 21. There appear in this table fully as many negative correlations as positive. The average of the entire ten is $-.016$, or practically zero. Two rather large coefficients are found, however — one of $-.45$ and one of $+.38$. In order to secure an empirical check on the significance of these values, ten rank coefficients were computed from chance drawings of 17 numbered blocks arranged in series comparable with the original data of Table 21. A coefficient of $+.47$ and one of $-.26$ were secured in this manner, the average of the ten coefficients in this case also being practically zero. It is accordingly quite evident that the coefficients shown in Table 21 represent nothing more than chance relations.

TABLE 21

CORRELATIONS BETWEEN CHARACTER AND HANDWRITING

	SPEARMAN RANK- DIFFERENCE COEFFICIENTS
Ambition with upward sloping lines	-.20
Pride with upward sloping lines	-.07
Bashfulness with fineness of line	-.45
Bashfulness with lateral narrowness of <i>m</i> 's and <i>n</i> 's	+.38
Forcefulness with heavy lines throughout . . .	-.17
Forcefulness with heavy bars on <i>t</i> 's	-.06
Forcefulness with heavy bars on <i>t</i> 's, corrected for size of writing	+.27
Perseverance with length of bars on <i>t</i> 's	+.00
Perseverance with length of bars on <i>t</i> 's, corrected for size of writing	+.16
Reserve with tendency to close <i>a</i> 's and <i>o</i> 's . .	-.02

And since they represent a fair sampling of the graphologists' claims, they probably may be taken with some assurance as typical of all such contentions.

A second investigation which was in part a repetition of that of Hull and Montgomery was performed by Lois E. Brown (8). Miss Brown began her investigation with a strong belief in graphology. She had as subjects 30 sorority women who ranked each other on five traits. Two of these traits were bashfulness and ambition, both traits which had previously shown relatively high correlations for Hull and Montgomery. Feeling that the copying of set material might not be a fair test of spontaneous writing, she secured letters written by the subjects when no thought of an experiment had existed. The writing contained in these letters was then measured much as in the preceding experiment, appropriate corrections being made where the writing paper differed from a standard size. Two of her handwriting traits, neatness and individuality, could not be measured objectively.

Accordingly, she had the samples of handwriting ranked for each of the two traits by ten independent judges. Pearson correlation coefficients computed from Miss Brown's data are shown in Table 22.

TABLE 22

SHOWING COEFFICIENTS OF CORRELATION BETWEEN TRAITS OF HANDWRITING AND VARIOUS TRAITS OF CHARACTER (Computations based on data of Lois E. Brown)

CHARACTER TRAIT	HANDWRITING TRAIT	CORRELATION COEFFICIENT (<i>r</i>)
Bashfulness	Width of down strokes	+ .11
Ambition	Tendency to upward slope as line crosses page	+ .23
Persistence	Width of down strokes	— .05
Persistence	Disconnected writing — per cent of breaks of line within words . . .	— .03
Personal neatness	Neatness in appearance of writing . .	+ .23
Personal individuality	Individuality in appearance of writing	+ .15

The first two coefficients from Brown's results represent practical repetitions of corresponding parts of Hull and Montgomery's investigation. The first coefficient of + .11 between bashfulness and width of down strokes corresponds to the Spearman coefficient in Table 21 of — .45 between bashfulness and fineness of line. Since the former is based on coarseness and the latter on fineness of line, the opposite-ness in sign in the two coefficients is really an evidence of agreement. Unfortunately, both are in direct opposition to the claims of the graphologists. Brown's correlation of + .23 between ambition and upward slope of line across the page is in conflict with Hull and Montgomery's coefficient of — .20 for the same variables. The two correlations relating to neatness and individuality are small but are both in

the direction which might be expected. It would not be surprising that these two last coefficients may turn out to represent genuine tendencies.

SOCIAL AGGRESSIVENESS AND CHEMICAL CONSTITUTION
OF URINE

It is inevitable that as aptitude science develops, not only anatomical but physiological and biochemical processes must be explored for clues which may contribute to the accuracy of aptitude forecasts. In the opinion of the writer the aptitude expert of the future must be much more than a psychologist. Among other things he will probably need to be conversant with certain branches of physiological chemistry. At present the work in this field has only just begun.

A suggestive investigation reported by Rich concerns the chemical constitution of the urine as an indicator of temperamental traits. He had academic fraternity brothers rate each other on five character traits. These ratings were correlated with the results of a quantitative chemical analysis of the urine for acidity, creatinin, and phosphorus. Rich reports the existence of a definite negative relationship between social aggressiveness and leadership, on the one hand, and acidity in the urine, on the other.

CHEMICAL CONSTITUENTS OF THE BLOOD AND CHARACTER
TRAITS

A second investigation reported by Rich¹ concerns the chemical constitution of the blood in relation to the general behavior traits of children who are being studied in a behavior clinic. Preliminary results indicate that blood phosphorus is a sign of good nature and that blood creatinin is a sign of excitability.

¹ Given the writer in a private communication.

A study reported by Starr grew out of an investigation of the relation of hydrogen ion concentration in the saliva to stammering. Encountering difficulties from irrelevant variable factors in the mouths of his subjects, he was led to the direct determination of the carbon dioxide tension of the blood. He reports that a common type of stammerer, also presenting a definitely lethargic type of general behavior, shows a great overload of carbon dioxide in the blood. Under treatment, in proportion as these patients showed a diminishing carbon dioxide tension in the blood they are also said to have become less lethargic. This suggests the desirability of investigating the same relations among normals. If it should lead to a useful chemical sign or index of "pep" or industriousness, it would satisfy a great need in our present supply of aptitude indicators.

An investigation carried out under the direction of Richard Van Tassel and the writer tested the redness of the blood in an attempt to secure a useful index of tendency to activity or "pep." It was thought possible that the amount of energy liberated in motor activity might be a function of the hæmoglobin in the blood, since this substance is known to carry the oxygen which is involved in the energy output. A rough but rapid colorometric test of the redness of the blood was made on 80 freshmen engineers at the University of Wisconsin. Later the results of this blood test were correlated with semester grades in English, mathematics, chemistry, mechanical drawing, and shop work. The coefficients resulting were all small, some being positive and some negative. The correlation with the average of the marks in all five courses was $-.02$. These results show pretty definitely that this test of blood redness is of no value as a sign of academic aptitude.

HEART RATE AND BLOOD PRESSURE AS SIGNS OF APTITUDE

In connection with the Van Tassel blood test just mentioned, the same 80 freshmen engineers were tested both for systolic and diastolic blood pressure and the pulse was counted for a full minute just before the beginning of a series of individual aptitude tests and again when the series of tests had been in progress about 40 minutes. Later these various scores were correlated with the grades obtained by the men at the end of the year. These correlations are shown in Table 23.

TABLE 23

SHOWING THE CORRELATION OF VARIOUS CIRCULATORY PHENOMENA WITH VARIOUS TYPES OF ACADEMIC SUCCESS ON 80 FRESHMEN ENGINEERS (Computed from Van Tassel's unpublished data)

ACADEMIC COURSE	CIRCULATORY PROCESSES			
	First Heart Rate	Second Heart Rate	Systolic Blood Pressure	Diastolic Blood Pressure
English	+.08	-.04	+.31	+.13
Mathematics	+.20	+.09	+.18	+.16
Chemistry	+.07	+.13	+.15	.00
Mechanical drawing	+.01	-.01	+.03	+.01
Shop	+.04	+.08	-.17	-.11
Average all courses	+.12	+.08	+.16	+.07

An examination of this table suggests that a high heart rate accompanied by high blood pressure preceding a psychological examination tends to be a sign that the subject will do well in his more intellectual courses, particularly mathematics. Motor activities such as drawing and shop courses seem not to be related, unless possibly shop work is correlated negatively with blood pressure. Of the four circulatory scores, the heart rate just preceding a psychologi-

cal test, and systolic blood pressure, seem to be most significant.

HEART RATE, BLOOD PRESSURE, AND EMOTIONAL UPSET DURING TESTING

Further computations from the Van Tassel data revealed a number of interesting relations between the circulatory processes just considered and certain tests. Among the tests which were given at the same time as those relating to the blood system were the following: the Thurstone technical information test, which is a multiple-choice group test containing 100 items; a complicated directions test relating to the opening of a puzzle box containing among other things a simple combination lock; and the miniature test for the engine lathe described on page 67, which gives two distinct scores — (a) the time required to pass around the circle of contacts and (b) the distance traversed. Both of these scores are to be as small as possible. The correlations between the four circulatory tests and the four voluntary behavior tests are shown in Table 24.

An examination of the correlations shown in Table 24 suggests rather strongly that high heart rate and blood pressure may have considerable influence upon test performance. In the case of the technical information test this is seen to be quite systematically detrimental, though in the directions test it appears to result in an advantage. In the case of coördination (speed) it results in slowing up the action, but with a compensating advantage in "distance." The apparently deleterious results on the information test which are correlated with the circulatory disturbances ordinarily associated with emotional upsets are of special interest. This interest arises in part because this is the type of test most widely used. Moreover, the negative correlation of these circulatory phenomena with this test performance

TABLE 24

SHOWING THE CORRELATION BETWEEN CIRCULATORY EVIDENCES OF EMOTIONAL UPSET AND VOLUNTARY BEHAVIOR UNDER TEST CONDITIONS
(Computed from Van Tassel's unpublished tables)

TEST PROCESS	CIRCULATORY PROCESS			
	First Heart Rate	Second Heart Rate	Systolic Blood Pressure	Diastolic Blood Pressure
Information test	-.30	-.33	-.18	-.19
Directions test	-.22	+.04	-.05	-.21
Coördination (speed) ¹	+.22	+.36	+.03	+.15
Coördination (distance) ¹	-.19	+.05	-.07	-.20
First heart rate		+.81	+.27	+.14
Second heart rate	+.81		+.22	+.21
Systolic blood pressure	+.27	+.22		+.34
Diastolic blood pressure	+.14	+.21	+.34	

attains special significance because the same phenomena are *positively* correlated with the more intellectual criteria such as mathematics.

It has been a matter of common observation ever since the beginning of the test movement that most persons, especially adults, regard a test more or less as an ordeal and that the consequent emotional disturbance during the taking of the test frequently distorts the test score. The above results suggest that possibly the pulse rate or the blood pressure (or both) may be employed to correct these distortions to a useful degree. Indeed, it would not be surprising that aptitude tests of the future should include pulse rate and blood pressure as regular components, largely for this purpose.

¹ Same as "Miniature test for engine lathe." (See Figure 11, page 68.)

CHAPTER FIVE

DERIVATION OF THE APTITUDE PROGNOSIS FROM RAW TEST RESULTS

A CRUDE test score as obtained directly from the measurement of test behavior is, as a rule, nearly or quite meaningless. Ordinarily a test score becomes of significance and value only when it is known to indicate with reasonable certainty the degree of a potential aptitude possessed by the person making the score. There is thus encountered in all actual aptitude work the problem of translating the raw test results, or other original evidence of aptitude, into a final aptitude prognosis. It will be convenient to divide our consideration of this fundamental problem into two parts:

1. The selection of the most desirable scale in terms of which to forecast a particular aptitude.
2. Methods and devices by which the scores of the tests of a battery may be combined and translated into equivalents of the aptitude scale.

DIFFICULTIES OF CHOOSING AN APTITUDE-PREDICTION SCALE

Superficially considered, the selection of the scale and unit in terms of which an aptitude shall be forecast is very simple. If conditions permit, the natural unit system for prediction is clearly the scale in which the aptitude itself is ordinarily measured. Thus, if scholastic aptitude is to be forecast, the natural scale should be that of the ordinary academic marking system. If the sprinting aptitude of an individual is being prognosticated, it would be natural to have it done in terms of the number of seconds it would require him to perform the 100-yard dash when at the limit of practice.

Unfortunately for the simplicity of the problem, many aptitudes have no recognized scale by which they may be measured, much less forecast. This is true despite the fact that the differences between the extremes in the aptitudes may be very great. As an example of this, we may consider the difference between the ability of Fritz Kreisler and that of the ordinary fiddler, in playing the violin. The difference here is sufficiently obvious; yet in what unit shall the difference be measured? For problems of this general type a variety of scales and units have been devised. We shall now consider the more important of these, together with their chief advantages and defects.

THE CRITICAL TEST SCORE

Perhaps the simplest method of interpreting test results in terms of aptitudes is on the basis of *critical scores*. This method is frequently used where individuals are to be divided into two sharply contrasted groups on the basis of test performance. A typical example of this is found in employment management, where an applicant is either to be hired or not hired. Here the critical score is that test performance above which candidates are regarded as satisfactory but below which all are rejected. In this case the attempt is made to find a *high* score above which few or none fail to make good on the job, even though a considerable number of good workmen may be turned away. A contrasting situation is found in scholastic-aptitude tests such as are administered to freshmen entering colleges and universities. In this latter case the attempt is made to locate a critical *low* score at a point below which practically all will fail to do a passing grade of work, and who may be denied entrance, even though many incapable students may be accepted. In the one case the stress is upon unmixed success; in the other, on unmixed failure. Despite these superficial differences, however, it

is evident that *the method of critical test scores is really based on a very coarse two-step scale of success and failure.*

As an example of the use of a critical score in employment there may be mentioned the work of Link on shell inspection. He tested with a number of different tests fifty-two women engaged in this work at an arms-manufacturing plant. Among the tests used was the number-group checking test. A large proportion of these women were efficient enough to work according to a piecework system, whereas the least efficient ones were paid by the day. The test performances of the two groups are shown in Figure 25. Link chose as his critical point the intersection of the lines outlining the respective distributions. This falls at 185 seconds. According to this all persons requiring less than 185 seconds to perform the number-group checking test would be eligible for employment as shell inspectors, whereas all those requiring more than 185 seconds would be rejected as unfit for the work.

As an example of the use of a critical score in scholastic aptitude, it is stated by the College Entrance Examination Board that, in general, of those freshmen making test scores

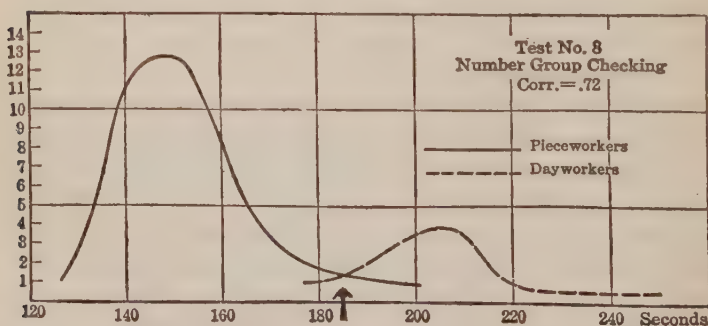


FIG. 25. Showing a graphic method of locating a critical score. For explanation, see text. (From Link's *Employment Psychology*. Reproduced by permission of the Macmillan Company.)

falling in the lowest 10 per cent of the entering class only two out of five will remain in college more than a year and only one will graduate. A second example comes from an investigation by Odell. He concluded from a study of the Otis Self-Administering Tests of Mental Ability given to 12,000 Illinois high-school seniors that an IQ of 90 on these tests is a lower critical score. He holds that those persons falling below this point have little chance of academic success. He also located an upper critical score at IQ 110 above which persons are likely to be successful in scholastic work.

Two factors detract from the value of the critical test score for use in aptitude prognosis. One is that correlations are never perfect. If, for example, college entrance tests should correlate perfectly with scholastic success, an exact test score might be found above which all students would be certain to succeed and below which all would be certain to fail. Under such ideal conditions a critical test score would correspond exactly to a critical aptitude score. As things are, however, the best that can be done in this direction is to say that where test scores fall below a given point the chances are such and such that the individuals in question will fall below a given point in the aptitude activity. If the critical score be lowered, the proportion of failures will be greater; if it be raised, the proportion will be decreased. Consequently it must always be a matter more or less of arbitrary judgment as to just what proportion of hits to misses should be fixed upon as most desirable.

Secondly, the critical point in the aptitude itself is subject to fluctuation. In employment, for example, the critical score may be shifted upward or downward according to the conditions of the labor market. If there is a plentiful supply of labor the score may be placed high, thus resulting in the acceptance of a small proportion of applicants but yielding a better quality. If, on the other hand, there is a

meager supply of labor the critical score may be lowered in order to secure enough workers, though at a sacrifice of including a somewhat larger proportion of unsatisfactory individuals. The critical score thus turns out to be not a point but a zone. In extreme situations this zone may involve the whole range of talent.

THE COARSE DESCRIPTIVE APTITUDE SCALE

We thus pass from the critical-score system, which reflects a very coarse two-step aptitude scale, to a system having several steps though still rather coarse. Such a system is represented by the army gradings on the Alpha test. The possible range in test score of this battery is from 0 to 212 points. On the basis of experimental comparison of test scores with actual army aptitude, the coarse practical scale shown in Table 25 was involved:

TABLE 25

SHOWING THE SCALE OF MILITARY APTITUDE EMPLOYED IN ARMY ALPHA
(Adapted from Yoakum and Yerkes in *Army Mental Tests*)

RAW TEST SCORES	COARSE APTITUDE SCALE		PER CENT OF RECRUITS
	Descriptive Terms	Letter Symbols	
135-212	Very superior	A	4%
105-134	Superior	B	9%
75-104	High average	C ⁺	17%
45- 74	Average	C	25%
25- 44	Low average	C ⁻	20%
15- 24	Inferior	D	15%
0- 14	Very inferior	D ⁻ or E	10%

The various specific aptitudes found to correspond to the respective steps of the scale are as follows:

"A" men are of high officer type when they are also endowed with leadership and the other necessary qualities.

"B" men are likely to be either of the commissioned or non-commissioned officer type.

"C+" men are usually suitable for non-commissioned officers, with occasionally a man who by virtue of a commanding personality is suitable for a commissioned officer.

"C" men make excellent privates, though some are good enough for non-commissioned officers.

"C-" men usually make good privates who do satisfactory work of a routine nature.

"D" men usually make fair privates, though they learn slowly, lack initiative, and require more than the usual amount of supervision.

"D-" men are barely fit for the regular service.

"E" men cannot make ordinary soldiers and should be recommended for development battalion or discharge.

PERCENTILES

A third device for mediating between crude test results and aptitude is percentiles. For practical purposes a percentile may be considered as the rank a person would take among 100 individuals in a perfectly normal distribution, counting from below upward. Thus a person at the 1-percentile is the poorest in a hundred, the person falling at the 10-percentile ranks tenth, and the person falling at the 100-percentile is the best in a hundred. A series of percentiles related to musical aptitude is given in Table 26. By this table it appears that the power of discriminating pitch as accurately as .25 of a vibration per second ranks best in a hundred, whereas a discrimination threshold of 34 is about the worst. A threshold of 3 is about median ability.

The percentile method of representing test results for the purpose of interpretation in terms of aptitudes has the advantage of being applicable to all kinds of test scores, no matter what the number of subjects in the experimental

TABLE 26

SHOWING THE PERCENTILES FOR PITCH DISCRIMINATION OF ADULTS
(After Seashore, *The Psychology of Musical Talent*, page 45)

THE JUST PERCEPTIBLE VIBRATION DIFFERENCE	PERCENTILE RANK
0.25 per sec.	100
1 " "	97
2 " "	80
3 " "	52
5 " "	32
8 " "	18
12 " "	10
18 " "	6
25 " "	3
34 " "	1

group may have been, or what the nature or unit system of the original test. It is thus a convenient method of reducing all kinds of test scores to a common mode of expression. In addition, the fundamental notion of ranks is easily grasped. For these reasons the percentile method has been used extensively, particularly among educational psychologists. Unfortunately, percentiles, like all systems of ranks, are not true linear scores. Their apparent simplicity is thus highly deceptive. The distances between consecutive steps near the middle tend to be much smaller than those between consecutive steps at the extremes of distributions. This is shown graphically in Figure 54 (page 384). The result of this distortion is that percentile values cannot be treated like the units of an ordinary scale. In percentiles, two added to two does not necessarily equal four. It is doubtful if percentiles are ever of value when used *directly* for purposes of prognosis in scientific aptitude work. They may, however, be used indirectly just as ordinary ranks may be. (See page 386.)

MENTAL AGE AND IQ

A fourth system of mediating between crude test results and the aptitudes which the tests are intended to throw light upon, is based on what is known as *mental age*. This was first proposed by Alfred Binet in the early days of mental testing. It was done in connection with the interpretation of the tests for children devised by himself and Dr. Simon, known as the Binet-Simon Tests. The unit in this system is a year of development in the power to respond to the tests included in the scale. In the Stanford revision of these tests, perhaps the most widely used of any, for most of the years of a child's life there are assigned six tests. For each one of these tests performed successfully, the child is credited with one sixth of a year of mental age — i.e., two months. The tests of the scale are of such a difficulty that the average child will be able to pass enough of them to make the mental age or credit secured in this way just equal to his actual or chronological age. If, however, a child of twelve years passes only enough tests to give him a score of nine years' mental age, the indication is that he is distinctly inferior in whatever is tested by the tests.

It has become common to divide the mental age by the chronological age (decimal point being dropped), with a view to securing a kind of index of brightness of children. This index is called the Intelligence Quotient, or more briefly the IQ. In the example instanced above,

$$\text{IQ} = \frac{9}{12} \text{ or } 75$$

This is generally taken to mean that the child in question has about 75 per cent of average intelligence. This index of brightness or general scholastic aptitude is supposed to remain approximately constant throughout the life of the individual. If, however, instead of passing tests enough to secure a mental-age score of nine years the child had passed

enough to secure a mental-age score of fifteen years, then the IQ would have been 125. The indication then would be that he is distinctly superior and will remain so throughout life.

The concept of mental age has served a most useful purpose. It was probably Binet's most brilliant stroke. The units were simple and easily comprehended by every one, which is a consideration of no small importance when a science is very young. Indeed, it is hardly possible that Binet's system of testing would have come into very general use without it. Nevertheless, with the increase of knowledge regarding tests it has gradually become clear that the employment of mental age as a unit in testing work involves great difficulties and some downright absurdities. The main trouble arises from the fact that mental ages do not mark equal steps of amount on a true scale. While the rate of mental growth throughout the early years of childhood is sufficiently constant to make up a scale for practical purposes, this is by no means the case as we pass up through the teens. After about twelve or thirteen years of age, each successive year thereafter sees a smaller and smaller amount of development, until at eighteen or twenty years of age mental growth in the ordinary sense apparently ceases almost altogether. It is therefore clearly impossible to equate the amount of mental growth during the tenth year with that during the twentieth year. The inadequacies of the system become most disturbing when the attempt is made to compute intelligence quotients from mental ages during these later years.

The situation may be illustrated from the field of physical growth. This field is chosen for purposes of illustration because there are no ambiguities arising from a doubtful unit of measurement and no inhibitions arising from metaphysical theories as to the nature of the mind. Figure 26 shows the growth in height of girls from the age of four years

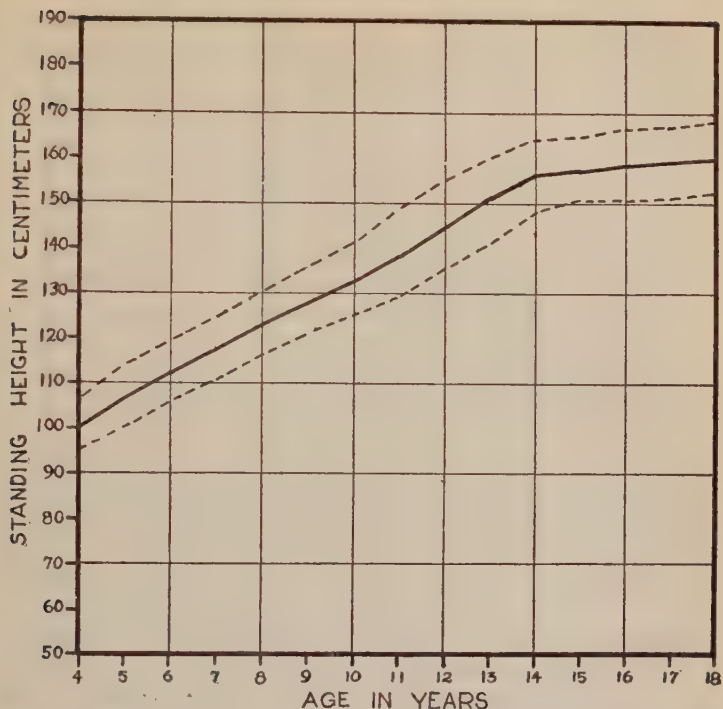


FIG. 26. Showing course of growth in standing height of girls. Heavy line shows median height for each age. The upper and lower dotted lines show the height at the 90-percentile and the 10-percentile, respectively. (Plotted from Smedley's percentile graphs as given by Whipple.)

to practical maturity, together with an indication of the range of height for each age. A glance is sufficient to show that the average amount of growth from year to year is by no means sufficiently constant for satisfactory use as a unit of measurement or expression. For example, the total amount of growth for the four years between fourteen and eighteen is less than half that for the one year between twelve and thirteen!

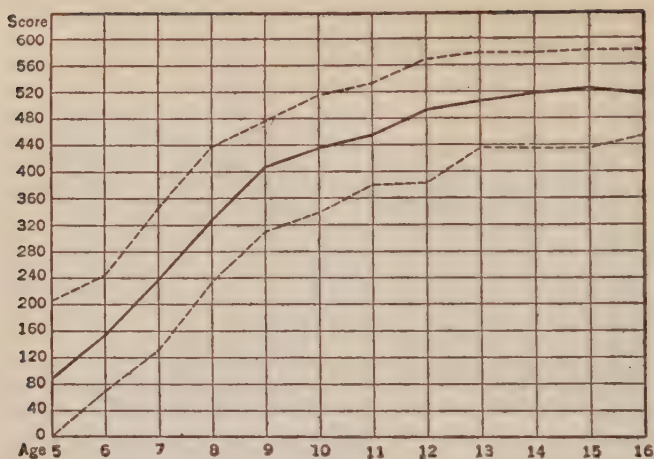


FIG. 27. Showing the increase of power to do the picture-completion test through the years of childhood. The solid line represents the median performance. The upper dotted line the 90-percentile and the lower dotted line the 10-percentile. (From Pintner and Paterson's *A Scale of Performance Tests*, page 132. Reproduced by permission of D. Appleton & Co.)

This same general situation is shown for test behavior, with which we are more immediately concerned, by Figure 27. This represents the growth in ability of children to do the picture-completion tests, together with an indication of the range of ability at each age. In this activity, maturity seems to have been reached at age fourteen, the rate of growth having diminished continuously for no less than five preceding years. At no year after nine does the amount of growth furnish a satisfactory unit for the expression of an aptitude — or indeed of anything.

SCALES BASED ON RANGE OF GROUP VARIABILITY

A fifth type of scale and unit of measurement is based on the range of group variability. A glance at Figures 26 and 27 will show that the *range* of performance at each age is

relatively constant even though the *amount of growth* from year to year is distinctly inconstant. The range in each case is indicated by the distance between the two dotted lines at any given age. In the case of growth in stature in girls it may be seen that during the early years, when growth is relatively constant, the amount of growth each year is maintained quite uniformly at about one third that of the distance between the 10- and 90-percentile. But after the age of fourteen, when growth is no longer at its original constant rate, the *range* is still maintained quite undisturbed. In this range of variability we find a relatively stable phenomenon which may serve as a satisfactory point of reference in defining universally valid units for the expression of test scores and aptitudes. If years of growth or mental ages be considered a vertical dimension, then the range or variability of the group at any given age may be considered the horizontal dimension. It is accordingly upon this stable horizontal range-of-talent or dimension that is based our fifth method of expressing test results for purposes of interpretation and aptitude prognosis.

There are several recognized methods of measuring the range or variability of groups. Of these, by all odds the best is the standard deviation. The general nature of the standard deviation has been described above (pages 28 ff.). Because of its excellence as an expression of variability of groups it has come to be accepted as the basic unit for measuring not only the horizontal or group-range dimension of test performance, but indirectly the vertical or growth dimension as well.

The point of reference in most physical measurements such as length, weight, etc., is a true zero, or "just not any" of whatever is being measured. Unfortunately many complex and abstruse considerations are involved in determining the true zero points for most forms of test behavior. Luckily

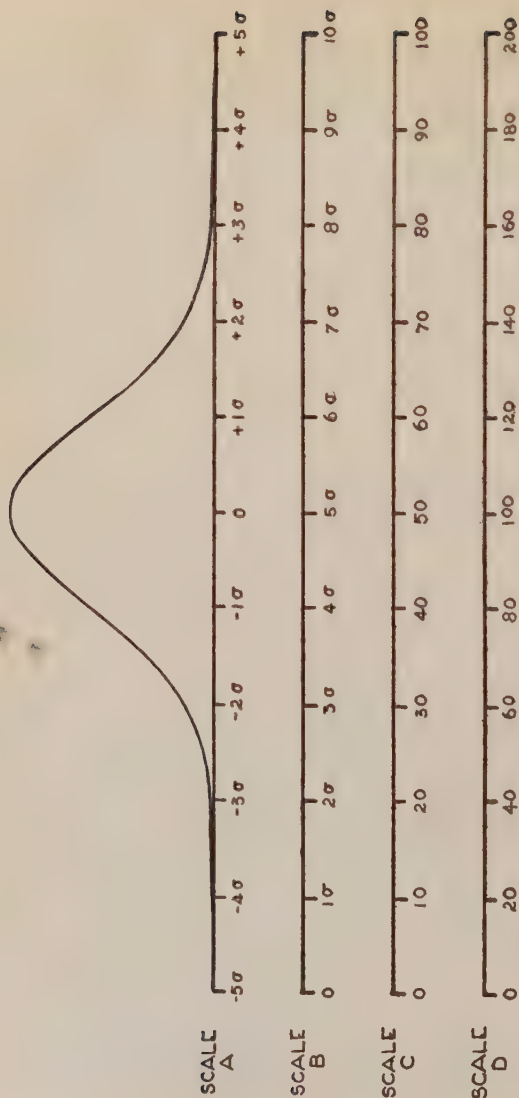


Fig. 28. Showing a variety of scales based upon the assumption of a "normal" distribution.

the location of the true zero point of a test is of more value in satisfying the scientific curiosity of psychologists than in practical usefulness. Accordingly, when constructing a scale on the basis of the S. D. (or σ , as it is more frequently called in this connection), the point of reference ordinarily taken is the *mean* of the group performance. But since the mean is in the middle of the group distribution, distances must naturally run in both directions. This introduces the complication of plus and minus signs. In order to distinguish the two directions, that one in which the variable increases is usually considered plus, whereas the direction toward zero is usually considered minus. This results in the scale shown at A, Figure 28. Although somewhat clumsy in use and not very intelligible to the uninitiated, it is reliable and has been used extensively by psychologists in technical research.

As is shown by Figure 28, the great mass of the individuals of the group fall within $3\frac{1}{2}\sigma$ of the mean in one or the other direction. Theoretically, however, a very few individuals may scatter out as far as 5σ , or even farther. Actually 5σ is the practical limit of the range of talent. An arbitrary or working zero point consequently may be placed at -5σ from the mean as the limit of normal variability. If we arbitrarily call this point zero, we now have a σ scale of ten points (Scale B, Fig. 28) exactly like the first except that the awkward + and - signs have been eliminated.

In case it is desired to have a finer scale than is provided by B, the σ may be divided into tenths and this $\frac{\sigma}{10}$ taken as the unit. Such a scale is shown at C. This makes an extremely neat 100-point scale. McCall has offered the ingenious proposal that such a tenth-sigma scale based on twelve-year-old children be made a kind of universal scale for use in testing work. The choice of twelve-year-old children as the basis for the construction of the scale was

very happy for several reasons. Among these may be mentioned the central location of twelve-year-olds in the growth process and the ease of securing relatively normal groups of such children in the schools for purposes of standardizing new tests. In honor of Thorndike and Terman, two American leaders in psychological testing, McCall has suggested that these units be called T's. Scores expressed by this system are consequently called T-scores. This method is rapidly coming into use in testing work with children. In respect to precision it seems distinctly superior to percentiles, mental ages, or any other system so far proposed; but it has the defect of requiring rather elaborate tables for its use.

A number of other systems based on the same general principles as that of McCall have been proposed and used to a certain extent. Otis, for example, has suggested for use an Index of Brightness, briefly indicated by the letters IB. This is a scale of 200 points, with its middle (100) located at the mean of the group. Its zero, however, is located at about $3\frac{1}{2}\sigma$ below the mean instead of 5σ as in Figure 28. But this is not a serious handicap, as less than one person in a thousand would fall below or above the limits of the Otis brightness scale.

The placing of the average performance of the group at 100 makes the IB somewhat similar in appearance to the IQ as used by Terman. The two values are not equivalent, however, since the zero point of Terman's IQ is much farther below the mean performance than even that of the T-score system. Terman states that half the population show IQ's within eight points of 100. This indicates that the zero in the Terman system must in practice be at approximately 8.43σ below the mean performance of any age group.¹

¹ Terman's statement is equivalent to saying that the P. E. of his IQ distribution is 8 IQ units. Since the σ is known to be 1.8426 times the P. E., the σ of Terman's distribution of IQ's must be about 8×1.8426 or 11.8606 IQ units. And since it is 100 units from his mean to his zero point, the zero point must be approximately $100 \div 11.8606$ or 8.43σ below the mean.

Aptitude work, even when children are being dealt with, is ordinarily looking forward to final adult potentiality. It is accordingly desirable that the ultimate goal of adult aptitude prognosis should have an adequate and standard form of expression. It is believed that the general principles of the scales already considered in Figure 28 are adequate for this service. It is therefore proposed that for the purpose of forecasting adult aptitude the point of reference shall be the mean *adult* performance. This will receive the value of 100 on the scale. The zero point will be placed as usual at 5σ below the mean. This makes a scale with total range of 200 points. (Scale D, Fig. 28.) It is suggested that the units of the scale be called A's, since they represent amounts of aptitudes. A score or aptitude forecast, when expressed in terms of this system, may be called an Aptitude Index, or, more briefly, AI. It may be observed that all but about one or two per cent of the population will fall between 50 and 150 on such a scale. It will have all the substantial virtues of IQ, yet will be based upon a more scientific foundation than the IQ's derived from mental ages.

THREE METHODS OF CONVERTING TEST SCORES INTO APTITUDE EQUIVALENTS

Aptitude tests can be of little value until the aptitude equivalent of the test score is known. The various devices by which the user of a test is enabled to find the aptitude equivalent of his test scores, therefore, deserves brief mention.

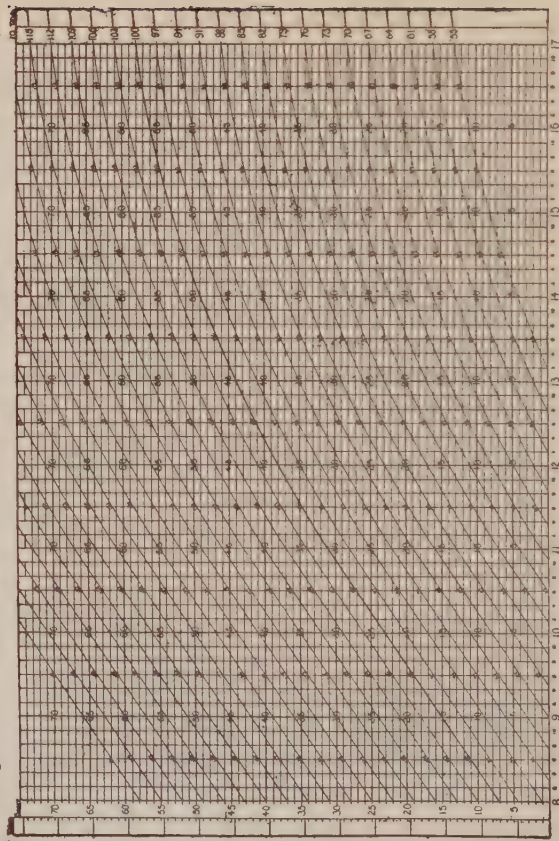
The most common method is to show the relation by means of a table. The test scores are arranged systematically in one column and the corresponding aptitude scores in a parallel column. An example of this is given in Table 25.

A second method of enabling the user of a test to find the aptitude equivalent of his crude test results is the use of line charts or nomographs. An example of this at its best is found

OTIS SELF-ADMINISTERING TESTS OF MENTAL ABILITY

INTERPRETATION CHART. For Intermediate Examination

Grades plotted.....Number.....Form used.....Time limit....min. Date....19...



Copyright 1922 by World Book Company, Yonkers-on-Hudson, New York, and 2126 Prairie Avenue, Chicago

FIG. 29.

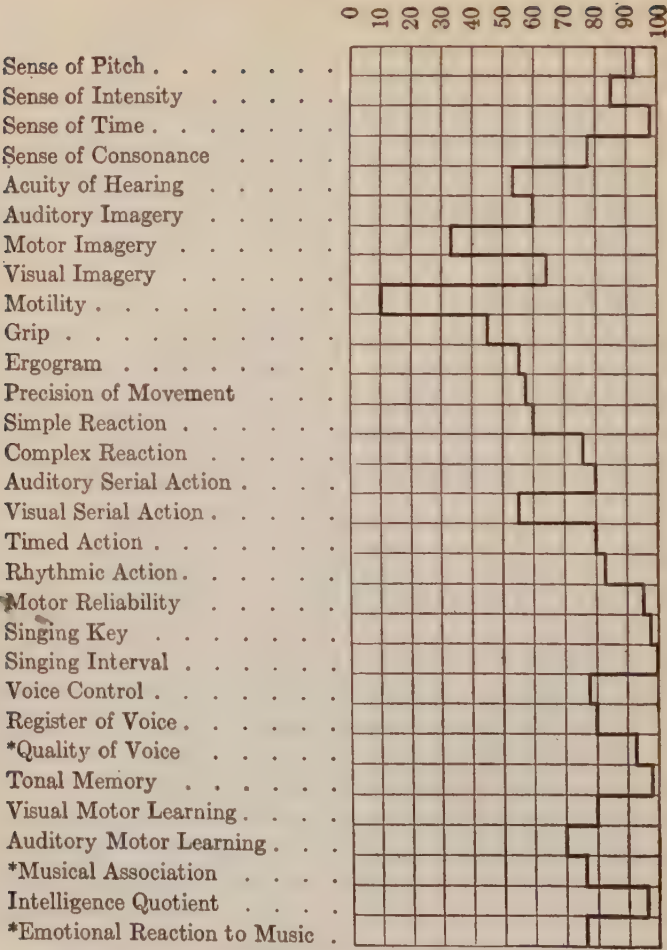
in Otis's Interpretation Chart for the Intermediate Examination of his Self-Administering Tests of Mental Ability. This chart is reproduced in Figure 29. The chart shows almost instantly the IQ of any test score for any age. Suppose a child of 14 years makes a test score of 55 points. Starting at the bottom on vertical line 14, you run up until you come to the horizontal line of score 55. Cutting through this point is a curved line numbered 106. This is the child's IQ.

A third method by which the user of a test may find the aptitude equivalent of crude test results is that of the multiple-regression equation and related forecasting formulæ. Since these formulæ also serve the function of combining the discrete scores of test batteries, it will be convenient to treat both matters together in a later section of the present chapter.

COMBINING TEST SCORES BY THE INDIVIDUAL PSYCHOGRAPH

Owing to the complexity of the psychological factors entering into most important aptitudes, it is rare for a single test unit to sample enough of the factors to secure a satisfactory prognosis. As a result, aptitude testing is practically always testing by means of batteries of tests. The number of units in a battery may vary from two or three up to a dozen or more. Seashore's phonograph tests of musical aptitude contain 5 units. Army Alpha has 8 units. Terman's Group Test of Mental Ability has 10 units. With the arrival of test batteries there has come the problem of how to bring the results of each of the test units simultaneously to bear upon the prediction of the aptitude under investigation.

Perhaps the most picturesque method of combining the scores of a test battery is that of the individual psychograph or profile. This is carefully to be distinguished from the job psychograph (page 294). The psychographic method has



Talent Chart of Theodora.¹

¹ The items marked with an asterisk in the charts represent mere estimates in the absence of norms.

FIG. 30. Showing a profile supposed to indicate a high order of musical talent. (From Seashore's *The Psychology of Musical Talent*. Reproduced by permission of Silver, Burdett & Co.)

evolved as a refinement of the clinical type of observation. Despite its sophisticated appearance, the psychograph is rather primitive in the sense that deductions made from it must be essentially qualitative in their nature. Typical psychographs of the better sort are shown in Figures 30 and 31.

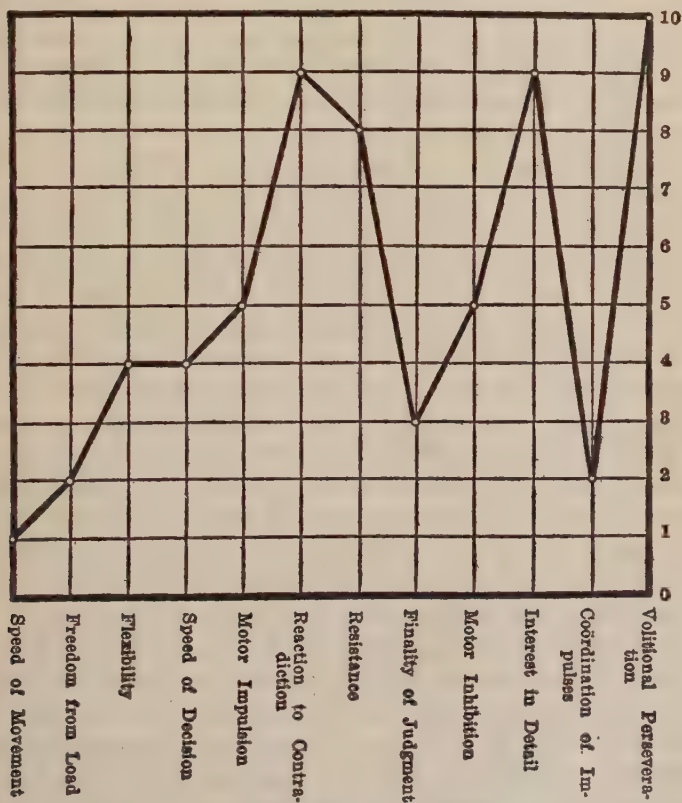


FIG. 31. Showing a temperament profile supposed to indicate a "slow, deliberated, accurate, highly controlled temperamental type." (From Downey's *The Will-Temperament and Its Testing*. World Book Company.)

Before a psychograph can be plotted, the crude scores of the various test units must first be converted into comparable measures. Such comparable measures in frequent use are percentiles, IQ's, mental ages, σ -scores, T-scores, etc. It should be observed that for this purpose, as has already been pointed out, percentiles suffer from the serious defect that they greatly exaggerate deviations near the mean (Fig. 54) while minimizing to an even greater extent deviations at the extremes of ability. The result of this is that psychographs or profiles plotted from percentiles present a distinctly distorted picture of the test situation.

The interpretation of a psychograph depends upon the skill of the prognosticator, very much as do the diagnosis and prognosis in clinical medicine. It is really a case of impressionistic judgment based upon the simultaneous contemplation of a multiplicity of data. The plotting of the data in the form of a graph is merely an optical device to permit the person making the interpretation to apprehend the data at a single glance. It appears to have been supposed by some that a special combination of high and low points on such a graph indicates special aptitude for one kind of activity, while a different combination of high and low points indicates special aptitude for something else, and so on. The general picture of the various types is supposed to be sufficiently characteristic to be recognizable by their general contours, much like the profile of an individual's face.

It is possible that for certain clinical purposes the psychological profile may be a useful aid in diagnosis. It is doubtful, however, whether it will ever be of much value in forecasting individual aptitudes, despite its superficial plausibility. In addition to the fact that the impressionistic judgments yielded by the method are necessarily rather general and indefinite, the whole method is based on a fundamentally false assumption. The implication is that it is possible for a

person without mathematical analysis but by some mysterious subconscious process (1) to determine the relative weights or degrees of importance that should be given to the various tests in order that in combination they should yield the most accurate prediction possible, and (2) without computation of any kind be able accurately to summate the various test scores in accordance with these weights. The human mind simply is not capable of performing such complex operations without arithmetical aids. It may be added that this general criticism holds in part concerning a number of the other methods presently to be described.

COMBINING TEST SCORES BY SIMPLE SUMMATION

A much simpler method of combining test scores is often used where the battery is made up of a considerable number of small and more or less discrete units which are scored separately. An extreme of this is found in the Binet-Simon Tests and certain group tests such as those of Otis and Miller, where the items, for the most part, are scored as either right or wrong. The method of combination is simply to count up the number of items correctly performed. A slight variation of the counting process of combining scores is also found in test batteries like Army Alpha. Here the separate test units are scored largely by the counting process, though in some units incorrect responses are subtracted from the correct ones. When it comes to combining the results of the various units of the battery, however, the totals of the various counting processes are simply added up in a column, to find the scores.

While no attempt is made by this counting method at especially weighting the various units making up a scale or battery, it may be said that the test scores are accurately summated at least, and every test unit contributes something to the final score. Not even this can be said for the psychographic or profile method. It may also be observed in this

connection that in test batteries where the various units are highly correlated, as in Army Alpha, the advantage in forecasting efficiency to be obtained from accurate weighting, while real, is not nearly so great as in the case of batteries where the various tests sample largely discrete types of behavior.

COMBINING TESTS ON BASIS OF ESTIMATED IMPORTANCE

A third method of combining the scores of battery units is to weight them in a rather simple manner according to supposedly shrewd guesses based on correlations with a more or less adequate aptitude criterion. The weights in such cases are usually very simple, some of the tests being added in just as scored, while others may be multiplied by a single small digit such as 2 or 3 or .5. The small numbers usually employed in such weightings have the advantage that the multiplication usually can be performed mentally in a moment by the person scoring the test at the conclusion of the scoring process. This matter of simplicity of operation is a factor of very great practical importance.

For the arbitrary method of weighting tests to be reasonably successful, however, much experience and skill on the part of the one doing the weighting is required. It is decidedly not a thing safely to be indulged in by amateurs. Indeed, it is about as likely as not that the forecasting efficiency of a battery thus arbitrarily weighted may be even less than at first. An arbitrary weighting should never be accepted without a careful check-up. Fortunately, the amount of loss in efficiency possible from a false weighting is not so great where the tests of a battery are highly correlated with each other as where they are relatively independent.

THE MULTIPLE-REGRESSION EQUATION AS A COMBINING
AND FORECASTING FORMULA

There is only one really adequate method of utilizing the full prognostic potency of a test battery. This method involves the use of a forecasting formula which is either a multiple-regression equation or is based on one.

The power of such prediction formulæ is quite amazing to those unacquainted with their use. These formulæ can, for example, convert test scores obtained in *minutes* and *inches* directly into an aptitude prediction in terms of *school grades*. Consider, for example, the miniature test of aptitude for learning the engine lathe, a sketch of which is shown on page 68. The subject must turn two small cranks simultaneously, one with each hand, in such a way as to move a point around a series of small brass contacts. The total distance the point travels in getting around the circle of contacts is measured in inches, and the time consumed is measured in fractions of a minute. It is desired to convert these test scores in inches and minutes into the most probable mark a subject will receive in a certain engineering shop course where the use of the lathe is taught. The formula for this particular forecast is :

$$\bar{X}_0 = 150.6 - 7.4 X_1 - 6.2 X_2$$

In this formula $\bar{X}_0$ is the predicted grade in the shop course, X_1 is the test score in inches, and X_2 is the test score in minutes. The coefficients of X_1 and X_2 are the weights of the respective scores, previously determined by scientific methods.

The making of forecasts by means of the formula is very simple. It may be illustrated by a concrete example. A young man by the name of Henry Wagner takes this test under standard test conditions. His scores are found to be 8.1 inches and 2.4 minutes. It is desired to forecast from

these test results, with as great accuracy as possible, the aptitude of the young man for learning to operate the lathe as taught in the shop course. His test scores are simply substituted in the formula, each is multiplied by its own weight, after which the various numbers are combined. The result is the most probable grade that can be estimated from these particular test scores:

$$\begin{aligned}\bar{X}_0 &= 150.6 - 7.4 \times 8.1 - 6.2 \times 2.4 \\ &= 150.6 - 59.9 - 14.9 \\ &= 150.6 - 74.8 \\ &= 75, \text{ or most probable grade}\end{aligned}$$

The indication is that Henry Wagner's mechanical aptitude is of a rather low order. Incidentally, minutes and inches have been transmuted into school grades.

The case just considered, while striking, is not typical, because instead of a test battery we have a single test which happens to yield two scores. We shall therefore examine a second aptitude prediction made by means of a forecasting formula, but one which is based upon a typical aptitude test battery. The battery is composed of five units, all group tests of the pencil-and-paper variety. The names of the tests, together with the corresponding symbols and a typical set of test scores made by a high-school freshman in a suburb of Milwaukee, follow:

NAME OF TEST	TEST SCORE OF DOROTHY SMITH	SYMBOL IN EQUATION
Best Answer (Terman) . . .	14	X_1
Sentence Meaning (Terman) .	10	X_2
Word Meaning (Terman) . .	0	X_3
Spelling (Hoke)	58	X_4
Multiplication (Courtis) . .	8	X_5

This battery was organized by Charles E. Limp for the purpose of forecasting aptitude to learn typing. The regression equation or forecasting formula derived for the battery is:

$$\bar{X}_0 = 73.4 - .23 X_1 - .36 X_2 - .16 X_3 + .26 X_4 + .45 X_5,$$

where $\bar{X}_0$ is the forecast of typing aptitude to be made and the remaining X 's are as indicated in the above table. The forecast is made exactly as in the case of lathe aptitude. The test scores are substituted in the formula and each is multiplied by its own weight, after which the resulting terms are united. This gives the desired forecast in terms of the most probable grade in typing likely to be received by Dorothy Smith at the end of the year:

$$\begin{aligned}\bar{X}_0 &= 73.4 - .23 \times 14 - .36 \times 10 - .16 \times 0 + .26 \times 58 + \\ &\quad .45 \times 8 \\ &= 73.4 - 3.22 - 3.60 - 0 + 15.08 + 3.60 \\ &= 92.08 - 6.82 \\ &= 85.3\end{aligned}$$

This forecast suggests that Dorothy Smith is gifted somewhat above the average in typing aptitude and should make a satisfactory typist.

The great advantage of the multiple-regression equation and forecasting formulæ of that general type is that they tend to give the best prediction of the desired aptitude possible to extract from the tests of any particular battery. This means that for the particular data from which the equation was derived any other conceivable system of weighting would estimate the criterion less accurately than the formula in question.¹ And it must be remembered that the tests of a battery are *always* weighted.² Test scores merely added together weight themselves according to the respective σ 's

¹ This assumes linearity of relation and additive method of combination.

² See page 457.

which they happen to show. Since accuracy of aptitude estimate is the prime reason for the existence of the test, this consideration is of decisive significance. A matter of such great importance surely cannot be left to mere chance.

There are, however, a number of disadvantages associated with the use of the multiple-regression equation in aptitude testing. One of these is that the deriving of the equation (see Chapter XIV) is a fairly complicated and technical procedure. It requires some time and labor of a specially trained worker. This is, however, but one more evidence that the work of devising aptitude tests is no longer a job for an amateur. The main thing is that any one can use the formula, once it has been derived.

A second difficulty hitherto associated with the multiple-regression equation as described above is that the arithmetic involved in combining the various test scores to secure the final estimate is somewhat more complicated and time-consuming than the simple addition frequently employed — e.g., in Army Alpha. The presence of terms some of which are positive and some negative in sign is especially awkward in casting up. Fortunately it is possible to eliminate from the use of the multiple-regression equation all the multiplication, all minus signs, and the manipulation of the constant term, by means of a small table which can be prepared in a few minutes. A table of this kind, together with an example of its use, is shown on page 485. In this connection it must be remembered that the multiple-regression equation not only combines the test scores scientifically, but that it also at the same time makes the forecast. By all other methods which yield at all detailed forecasts the test scores, after being combined, must also be looked up in a table, nomograph, or some similar device, before the corresponding aptitude can be determined. By the method described above for the use of the multiple-regression equation the table is merely

used at an earlier stage of the process. It is thus apparent that even in the matter of convenience of use for deriving the aptitude prognosis from raw test scores, the multiple-regression equation is very nearly the equal of the cruder methods. When its greater power of bringing the real potency of the tests to bear upon the aptitude is considered, it becomes clear that it is by far the best method of making the aptitude prognosis. For this reason it will be presented in full detail in Chapter XIV.

CHAPTER SIX

THE BASIC CONSTITUTION OF APTITUDES AND TESTS

ONE of the most fascinating problems in the world is that concerning the ultimate causes of human success and failure. For purposes of convenience, the gross determiners of success may be divided into (1) those which are accidental or chance factors — i.e., those lying quite outside the individual — and (2) those factors lying within the individual himself, of whatsoever source. Each of these groups of factors may be considered from a variety of points of view.

CAPACITY *vs.* INDUSTRY AS GROSS DETERMINERS OF SUCCESS

One of the most significant contrasts among the factors residing within the individual is that between *capacity* and *industry*. Various opinions have been held as to their relative importance in contributing to achievement. In some quarters it is common to emphasize the importance of industry. In this vein genius is defined as an unusual capacity for hard work. In somewhat more pungent terms it is asserted that success is made up of "one-tenth inspiration and nine-tenths perspiration." This one-sided emphasis is not difficult to understand quite apart from any truth the statement may contain. When a person is striving to achieve something, whatever special capacity he may have takes care of itself. The *capacity* requires no thought on the part of the striving individual. His attention is absorbed, instead, by the effort he is putting forth. When, therefore, a successful person is asked the secret of his success, he naturally finds little in his memory of the period preceding his success, except the effort. This, coupled with a wholly natural modesty, is likely to lead him to minimize any special capacity that may have existed and to assert stoutly that any one else with an

equal amount of effort could have done as well. It is probable also that inspirational writers and speakers have contributed much to the propagation of this view. Since the natural capacities of those they seek to uplift cannot be expected to yield to their efforts, they quite naturally turn to the volitional side and stress the importance of industry, which may really be stimulated.

Somewhat in contrast to the above is the tendency in recent years on the part of certain psychologists engaged in intelligence testing. The tests during this period have been of value chiefly as indicating capacity and of little or no value as indicating potential industry or tendency to put forth effort. It is therefore quite natural that with certain psychologists special stress should have been placed on capacity or natural endowment and sometimes a tendency shown to lose sight of the importance of industry in contributing to achievement.

ILLUSTRATIVE CASES

But probably all would agree that for any high degree of success to be attained, both capacity and industry must be present in liberal though not necessarily equal amounts. The most persistent effort will accomplish nothing of consequence without at least moderate capacity. No more will the greatest natural capacity accomplish anything of importance unless accompanied by a certain amount of industry. Striking examples illustrating these statements are occasionally met with. Take the case of Archie D. This feeble, shambling, apologetic little fellow, having somehow secured enough education to teach a school in a remote rural community, conceived an undying ambition to graduate from college and enter a learned profession. At length he found his way to a small college where feeble capacities were not looked upon too severely. Then followed five years of the most persistent application imaginable. At the beginning of the sixth year

the president of the college was forced to tell him that he could never hope to graduate. After that the college saw his timid, peering, mouse-like face no more. When last heard from he was washing glassware in a chemical laboratory.

By way of contrast may be cited the case of Frank E. This handsome, intelligent young man, slightly tainted by heredity, had from boyhood been troubled by a disinclination to put forth persistent effort. During high school and college he got along fairly well by occasional spurts of effort in academic emergencies. During these brief periods he demonstrated that he had capacity of exceptionally high order. He even passed successfully the examinations for the first year of graduate study. After much prodding he finally completed the experimental work for his M.A. thesis. But when it came to working up the data and completing the thesis itself, he stuck fast. He stayed on during the following summer, but the thesis was not finished. He returned the next September to concentrate all his energies on the thesis. At the end of the year, during which nothing else was done, the thesis was hardly started. He remained through the following summer and returned once more in September. During this second year also no other work of any kind was attempted. He was encouraged, prodded, and threatened in turn. Finally, after two full years of sincere daily resolution to work at the thesis, yet rarely doing so, he was induced to show what he had accomplished. He exhibited a series of computations which could easily have been performed in a week of ordinary application.

RELATIVE IMPORTANCE OF CAPACITY AND INDUSTRY AS
GROSS DETERMINERS OF SUCCESS

The spectacle of such extreme asymmetries of development is indeed painful. Fortunately, in the degree represented by the above examples, they are comparatively rare. And while

they illustrate very well the need of both capacity and industry for success, they give us no light on the relative importance of the two factors when combined in amounts within the ordinary range. There is, however, a certain amount of scientific evidence bearing on this question, although, like so much of such evidence, it comes from the rather narrow field of scholastic achievement. T. L. Kelley (44) had teachers rate first-year high-school students on (1) intelligence, (2) industry, and (3) interest in their work. He then sought to determine the relative part contributed by intelligence (capacity) and industry in determining scholastic success as measured by the average mark at the end of the year. He finally concluded, after considerable statistical analysis, that intelligence and industry contribute to scholastic success about in the proportion of 60 to 40 respectively.

A second investigation bearing on the same point was carried out by Mark A. May (57), at Syracuse University. He gave the freshman class a very elaborate series of the best available group tests of scholastic intelligence or mental alertness. The scores from these tests were taken as a measure of scholastic capacity. In addition he had them each fill out a questionnaire stating how much time was spent per week on each of various activities. Among these was the amount of time spent in study. The answer to this latter question was the real point of the questionnaire and was taken as the best procurable measure of scholastic industry. A figure derived from all their university grades was taken as a measure of academic success. With these data it was possible to secure some indication as to the relative importance of capacity and industry in determining the academic success of university students. Computations from May's published results indicate that capacity and industry should receive weights in the proportion of 58 to 42 when estimating academic success. These results accordingly support in an interesting

manner the findings of Kelley published some ten years previously. The obviously imperfect methods necessarily employed by both investigations in measuring the variables, as well as some uncertainty as to the exact significance of the weights¹ found, make us cautious in drawing conclusions. Still, their general harmony with ordinary observation as well as with each other indicates that they probably are not far from the truth, at least with respect to scholastic success. It would not be surprising, however, if future investigation should reveal many vocations or lines of activity in which the preponderance swings decidedly in the direction of industry at the expense of intelligence or capacity.

ARE THE ABLE OR THE POORLY ENDOWED THE MORE
INDUSTRIOUS?

The question as to whether a person of superior capacity is likely to display more or less than average industry is of considerable importance. In Kelley's study mentioned above the two were found positively correlated to the extent of .61. This means that upon the whole children estimated by their teachers as of superior intelligence were rather generally estimated to possess superior industry. May reports, on the other hand, a *negative* correlation ($-.35$) between intelligence-test score and number of hours spent in study per week. This tendency has frequently been observed by others. May's negative correlation means that on the whole the more intelligent university students studied less than the more poorly endowed, though there are many exceptions. The two investigations thus appear superficially to be in conflict. The high positive correlation of Kelley may be accounted for in part at least by assuming that his teachers were not able to distinguish clearly between capacity and industry, since both obviously combine to produce the same school behavior. May's results, on the other hand, are free

¹ See note, page 278.

from the difficulty just mentioned and doubtless represent the facts of his situation rather accurately. Indeed, it is well known that poorly endowed students are likely to study more than well-endowed ones. This naturally results from the pressure put upon the poorly endowed to maintain the minimum standard of performance. But since there is little in most life situations that resembles a "passing mark," the negative correlation found by May loses much of its generality. The most general statement that can be made in the light of the available evidence is that the well-endowed are probably a little more likely to be industrious than the poorly endowed, but not much.

Intimately connected with industry and success is the interest in or the liking for a given activity. Indeed, liking for a given kind of activity undoubtedly determines in large part the amount of industry and success displayed in it. It is notorious that people find time to do the things they like to do and discover an astonishing variety of excellent reasons for not doing the things which they dislike. In harmony with the results of general observation, Kelley found, in the investigation previously referred to, that the interest of high-school students in their work as judged by teachers was correlated to the extent of $+.68$ with industry or conscientiousness. This indicates a very considerable similarity, as such relations go, though by no means a complete identity. When the two factors were definitely compared as to their importance in determining (estimating) average scholastic success, Kelley found that the interest factor was only about half as important as industry.

CORRELATION OF THE INTERESTS OF THE INDIVIDUAL WITH HIS ABILITIES

Kelley was concerned with *general* interest and *average* success. In contrast with this, Thorndike (89) attempted to

determine how far the degree of a person's special interest in each of a variety of school subjects was associated with a corresponding degree of academic success in the same subject. He had 100 third-year college students rank themselves as to their interest in mathematics, history, literature, science, music, drawing, and handwork such as carpentry, sewing, etc. The ranking was done not only for the college period, but also (separately) for the high-school period and for the later elementary grades, as these were remembered. At a later date he had them rank their abilities as estimated on the same seven subjects for each of the same three periods. Some years later the experiment was repeated in detail with 344 students, with almost identical results (87). He found that in general the order of a person's interests for any given period correlated $+ .89$ with the order of the same person's estimates of his abilities for the corresponding period. But when the order of interests in the elementary school was compared with the order of estimated abilities in the college, some eight or nine years later, the correlation fell to $+ .66$. Thorndike, while recognizing certain weaknesses in his data, believed that they showed the importance of early interests as symptomatic of later success.

In order to test the validity of Thorndike's assumptions, Bridges and Dollinger (7) repeated the above experiment, but with an additional feature. They found interests correlating with estimated abilities to the extent of $+ .57$, which is high but somewhat lower than found by Thorndike. At the end of the semester, however, they correlated the interests with the grades actually received by the students. The correlation obtained was only $+ .25$. It is interesting also to note that the estimated abilities correlated only $+ .28$ with their actual abilities as shown by the grades received, thus showing a markedly feeble capacity of the subjects accurately to rate their own powers. This investigation

seems to indicate that the relation between interest and achievement, while positive and real, is distinctly less than Thorndike supposed it to be.

Along the same line of inquiry, Franklin (20) attempted to discover whether interest, as indicated by the choice of vocation, is associated with any special capacity in the type of activity chosen. He tested 635 7-B pupils in a junior high school with two standard group intelligence tests. Of these children 135 had expressed a choice for a clerical career. Computation from Franklin's published results shows that the 135 clerical students gave an average intelligence quotient of 99.3, whereas the remaining 500 of the original 635 gave an average intelligence quotient of 106.5. The clerical students, accordingly, rate distinctly below the remainder of the group in the mentality measured by these tests. Next Franklin gave to the same students the Columbia Institute of Educational Research clerical test, which is designed to give a special prognosis of clerical aptitude. In this test the 500-group averaged 34.8 points, whereas the 135 clerical reversed the score they received on the intelligence test, making the very high average of 47.5. Franklin accordingly concludes that interest is undoubtedly a potent factor in aptitude for success.

IMPORTANCE OF FORTUITOUS CIRCUMSTANCES AS GROSS DETERMINERS OF SUCCESS

Turning now to the accidental or chance factors, we shall consider the rôle of those causes lying quite outside the individual himself. There is no question but that chance factors, such as special opportunities on the one hand and special ill fortune on the other, play a genuine part in determining the degree of individual success or failure. It is also evident even to ordinary observation that the importance of fortuitous circumstances varies greatly in the lives of different individ-

uals as well as in different types of activity. Unfortunately, owing to the absence of precise evidence on the point, we can at present only conjecture the quantitative importance of these factors. But since a well-considered guess may have some value even if it serves no other purpose than to raise the question in a specific manner, one will be ventured for what it is worth. The writer's estimate is that, upon the whole, purely chance factors contribute to the aggregate of the determiners of individual success somewhere between ten and twenty per cent.

Owing to the inherently unpredictable nature of chance occurrences, it may seem at first sight that the behavior resulting from this, the fortuitous component of the determiners of success, might be forever hidden from the psychologist who would forecast individual achievement. This is true only in part. Fortunately a portion of the chance happenings in question antedate by months and years both the prognosis and the activity whose success or failure is being predicted. These antecedent, chance determiners of individual success are thus likely to be known and may therefore be taken into consideration. This may be done in two fairly distinct ways.

In the first case the fortuitous circumstance may be known to the psychologist by verbal report of the subject or otherwise and may be utilized directly in making the prognosis. An obvious example of this is the case of a subject who has lost sight or hearing through an accident, which fact alone would reduce to the vanishing point the possibility of success in certain forms of activity. A very different example is offered by the subject who has been badly spoiled as a child, though the prognostic significance of such a circumstance for any given kind of success might be difficult to estimate. Other chance factors which might be mentioned are the possession of wealth through inheritance, influential relatives,

the type of educational institutions attended, etc. As a final example may be mentioned the fact, recently discovered, that married men with one or two children are likely to sell more life insurance than unmarried men or married men whose wives are employed.

In the second case the fortuitous circumstance, while not known to the psychologist directly, may influence in a significant manner the subject's behavior in an interview or in his responses to various psychological tests, and in this way aid in the prognosis. One of the most obvious examples of this is the well-known manner in which the nature of previous educational opportunities influences the performance of subjects on ordinary intelligence tests. Also, most individuals retain powerful emotional tendencies, prejudices, preferences, etc., which have been produced by chance experiences. No doubt in time tests will be devised which will reveal more or less completely such tendencies, even though the circumstances giving rise to them may have been quite forgotten by the subject.

In conclusion, we may summarize the foregoing evidence as to the relative importance for success of the determining factors considered above. Assuming each to be disentangled from the complex overlappings of the others, their respective contributions are judged to be approximately as follows :

Capacity or ability	50 %
Industry or willingness	35 %
Chance or accident	15 %

The above quantitative estimate is made with full appreciation of the probability that as scientific evidence accumulates in this field, the percentages will require revision, possibly to a very considerable degree.

PROBLEM OF GENERALIZATION *vs.* SPECIALIZATION OF TRAITS

In considering the gross determiners of success, ability and industry have been treated as unanalyzed and undifferentiated wholes. In this way we have obtained a preliminary survey of the human factors involved in aptitudes. But before any very satisfactory advance into the essential psychology of special aptitudes can be made, we must seek to understand more fully the inner structure and organization of the behavior upon which aptitudes are based.

One great question to be considered in this connection is the extent, if any, to which various traits are generalized. Is a man equally industrious in all things, or does he have different amounts of industry for different kinds of activity? Is he equally industrious or lazy in looking after business matters, playing bridge, cleaning out the furnace, helping his wife purchase millinery, fishing for brook trout, soliciting funds for charity, and so on? Are people honest in general, or are they scrupulous in some things while being just ordinarily honest in others? Are people generally aggressive, or are some people mainly aggressive with respect to persons, others mainly aggressive toward things, and still others mainly aggressive toward theoretical problems or ideas? Does the individual show an equal amount of neatness, generosity, courtesy, patience, speed of decision, or courage in all situations? Does he discriminate the meanings of words with the same precision as he discriminates the pitch of tones or the merits of business proposals? Is his power of analysis the same for spatial patterns, logical fallacies, human motives, and problems in machine design? These and hundreds of similar problems which might be mentioned are not mere matters of academic interest. On the contrary they are fraught with the greatest significance in connection with the methods, possibilities, and future of aptitude prognosis.

The final answer to many, if not all, of the questions raised above must await further investigation. Many of them have received little and some practically no attention from scientific psychologists. This is particularly true of moral and character traits because of the difficulty of carrying out objective experiments on them. Accordingly, in those fields where experimental evidence is largely lacking we must for the present judge the situation as best we can on the basis of observation. The verdict by this method would seem to be that moral and character traits are neither wholly specialized nor wholly generalized. There appears to be a certain tendency to generalization, but in most cases it is probably not very strong. The chances are that in personal industry, for example, the various potentialities of this trait possessed by any given individual in a hundred of his possible activities, chosen at random, will show an appreciable tendency to cluster in some one part of the possible range, but that despite this tendency the scattering will be very wide (Fig. 10).

SPEARMAN'S THEORY OF GENERAL AND SPECIFIC FACTORS
AS APTITUDE DETERMINERS

Fortunately, in the extremely important field of "intelligence" a great deal of careful work has been done. The problem was first attacked by C. Spearman (79), an English psychologist, who published in 1904 an epoch-making article on the subject. He showed what has now become commonplace, that all kinds of intelligence tests tend to be positively correlated with each other. In fact, genuine negative correlations in this field are almost unknown. In addition he called attention to the fact that if all the possible correlations among a set of intelligence tests be calculated and then certain of these coefficients be picked out and arranged in a series of decreasing size, all the other coefficients when arranged by lists in the same order as the first, tend also to fall

TABLE 27

SHOWING THE FAMOUS HIERARCHICAL ORDER¹ OF SPEARMAN INTO WHICH CORRELATION COEFFICIENTS MAY OFTEN BE ARRANGED (Adapted from Spearman)

Note that as one moves down any column or to the right on any row the coefficients grow smaller with considerable regularity.

	CLASSICS	FRENCH	ENGLISH	MATHE- MATICS	DISCRIM- INATION	MUSIC
Classics . . .		+.83	+.78	+.70	+.66	+.63
French . . .	+.83		+.67	+.67	+.65	+.57
English . . .	+.78	+.67		+.64	+.54	+.51
Mathematics .	+.70	+.67	+.64		+.45	+.51
Discrimination	+.66	+.65	+.54	+.45		+.40
Music . . .	+.63	+.57	+.51	+.51	+.40	

into series of decreasing size. It is perfectly obvious that some causal factor must be at work here, since chance alone would not produce such a situation. Interpreting these results by means of certain ingenious and subsequently fruitful mathematical formulæ, Spearman concluded that intelligence is generalized to a rather high degree. He believed that the positive and fairly high correlation found among all intelligence tests was due to the presence of a single common factor which he was inclined to call "general intelligence."

According to Spearman's original view, a person's success in any given activity was determined by two factors or sets of determiners. There was in the first place the general or "*g*" factor which was supposed to be present in all intellectual activities. This is indicated by factor I of Table 28. In addition there was supposed to be a group of special or specific factors (*s*), which, jointly with the general factor, determined

¹ It should be stated that the hierarchical order as evidence of the two-factor theory was only a makeshift. More recently a far more precise method known as "tetrad differences" has been devised by Dr. Holzinger and used extensively by Spearman. There is, however, no essential difference in the basic logic of the two criteria.

TABLE 28

SHOWING THE ORGANIZATION OF APTITUDES ACCORDING TO AN EXTREME
"TWO-FACTOR THEORY"

Factor I is a general factor, since it is a component in all aptitudes, whereas the others are specific factors, since each appears in only a single aptitude. The amount of aptitude "A" possessed by a given person is determined by the amount of general factor I together with specific factors III and V.

APTI- TUDE	FACTORS OR DETERMINERS															
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
A	×		×		×											
B	×			×				×								×
C	×										×			×		
D	×						×		×			×			×	
E	×	×				×										
F	×									×			×			
etc.																

the degree of success in the activity. These specific factors were at first regarded as very special indeed in that they would never be found at all except in some other activity almost identical in nature. They were thus supposed to be all but unique. They are represented by determiners II to XVI of Table 28.

APTITUDES AS DETERMINED BY SPECIFIC FACTORS

Spearman's work at once gave rise to attempts to test his theories by further experimentation. Some experimenters reported results apparently confirming Spearman's contentions, others results which were regarded as indicating that no such thing as a generalized intelligence exists. Among the latter, E. L. Thorndike was much more impressed by the elements of uniqueness found in the various tests than by the indications of a general factor. The extreme form of this contrasted theory is illustrated by Table 29. Spearman,

TABLE 29

SHOWING THE ORGANIZATION OF APTITUDES ACCORDING TO THE EXTREME
SPECIFIC FACTORY THEORY

The determiners are all specific, none appearing in more than a single aptitude.

APTITUDE	FACTORS OR DETERMINERS															
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
A	×		×		×											
B				×				×								×
C											×			×		
D							×		×			×			×	
E		×				×										
F										×			×			
etc.																

however, insisted that the results of all experiments alike were completely in harmony with his theory of two factors, provided they were first treated according to his special mathematical methods.

APTITUDES AS DETERMINED BY A COMBINATION OF GROUP AND SPECIFIC FACTORS

At this juncture Godfrey Thomson appeared on the scene (9) with a destructive criticism of Spearman's mathematical methods of treating his empirical data so as to bring them into apparent harmony with his theory. Thomson was able to show by means of certain ingenious dice-throwing experiments that hierarchies of correlation coefficients similar to those taken by Spearman as evidence for a general factor can be produced when a general or common factor is known not to be physically present.¹ In these dice-throwing experi-

¹ The common or general factor was absent in the sense that it is absent in Table 30; i.e., certain factors were common to several of the variables, but no factor was a physical component of all as may be seen by inspection.

resulted from a general and a specific factor. But the physical fact is that in Thomson's dice-throwing experiments the factors were *not* so divided. "Nevertheless the tetrad equation was almost perfectly satisfied." This is frankly admitted by Spearman (80). Garnett's paper thus amounts to a general mathematical proof that all of Spearman's criteria adduced as evidence of the two-factor theory are satisfied by numerous overlapping group factors. This is, of course, exactly Thomson's contention.¹

It must be emphasized that Spearman recognizes, in addition to the general intellectual factor g , several other general or wide-range group factors of a non-intellectual nature. One of these is called c and is thought of as quickness, originality, or cleverness. A third general factor which is regarded as having been experimentally isolated is the tendency to oscillate or fluctuate in the general level of behavior efficiency. A fourth general factor, called w , is thought of as the persistence of motives resulting from volition or will. Lastly, there is a broad group factor which is supposed to appear in mechanical ability and is regarded by Spearman as related to the "instinct to play with mechanical toys."

¹ Just as the present work was going to press the author had the privilege, through the courtesy of Dr. K. J. Holzinger, of reading in manuscript an important paper by Stuart C. Dodd. Dr. Dodd shows mathematically that a strict group-factor theory, substantially like that put forward in the following pages, is rigorously consistent with the known facts of test and aptitude behavior. It may be added that, like the present writer, Dodd also abandons the specific factors. He finds that both Thomson and Spearman have been essentially right in their contentions, but each in a different sense. The present indication seems to be that a strict group-factor theory fits the actual physical determiners of aptitude and test abilities. These are probably what most psychologists think of as factors or determiners. But Spearman's central factor also appears to exist, possibly as a kind of mathematical expression of the totality of all the group factors. A central factor in the sense of Table 28 is clearly no longer held by the Spearman school, if indeed it ever has been. Thus a final agreement seems gradually approaching.

IMPLICATIONS OF THE VARIOUS THEORIES OF APTITUDE
DETERMINATION

We may now pause to note the implications of the various theories of the nature of aptitude determination (assuming each in turn to be true) upon the possibilities of aptitude testing. It will be convenient to begin with the extreme theory which recognizes only specific factors. Pressed to its logical limit, this specific-factor theory would hold that the factors determining success in all activities — *both aptitude behavior and test behavior* — can be found in no other activity whatever; i.e., that they must be absolutely unique. Thus a test activity could never contain any of the determiners found in an aptitude activity. There could therefore never be a correlation between the two. It is clear that such a situation would preclude any aptitude testing in the ordinary meaning of the term. There would remain, however, the possibility of utilizing a sample of the entire aptitude activity itself as a means of determining aptitudes. This would involve a kind of trial apprenticeship system. It is interesting to note that such a system has received a certain amount of favorable consideration.

We pass next to the aptitude implications of the theory of combined general and specific factors, assuming it, in turn, to be pressed to its logical limit. Because of the presence of the general factor in all activities (both in aptitude behavior and test behavior), tests would *always* correlate with aptitudes. Indeed, it would be impossible to find tests which would *not* correlate with *every* aptitude. Because of this it should be relatively an easy matter to secure tests which would give a fair indication of the general level of a person's ability as dependent on this general factor. Such a test should be useful in employment work, where the main interest is usually in discovering whether one person is more promising than

another for a given job. The more important the general factor as a component determiner of success in a given vocation, the more valuable such a test would be. Conversely, the larger the number of specific factors involved in an aptitude, the less valuable such a test would be. This would be true for the same reasons that explain why a strict specific-factor theory would preclude all testing in the ordinary sense.

It is interesting to note that since, according to this theory, each person has a certain quantum of this general factor, he will be equally effective in all aptitudes in so far as they depend upon this factor. It follows that any differences which a person may show in the efficiency of his various aptitudes must be due to specific factors. But inasmuch as the very specificity of such factors must preclude their being reached by tests, the two-factor theory would preclude the possibility of forecasting by means of ordinary tests *whether a person would be more effective in one aptitude than another*. It would be impossible to differentiate the aptitudes within an individual. On this theory, a vocational guidance based on psychological testing would be impossible. There would of course remain, just as in the case of the strict specific-factor theory, the possibility of determining the differences among the aptitudes of an individual by means of try-outs in the activities or vocations in question, either actually or by means of elaborate miniature tests.

Turning now to the Thomson theory of group and specific factors, we find it clearly offers greater possibilities in the way of aptitude testing than does either of the two theories already considered. This is due to the assumed presence of group factors, but is in spite of the specific factors. The implications of group factors for aptitude testing will be taken up in connection with the following theory of aptitude determiners upon which the present work is based.

A STRICT GROUP-FACTOR THEORY OF APTITUDE
DETERMINATION

The writer's view of the constitution of aptitudes and tests may best be characterized as a strict group-factor theory. It agrees with Thomson in not assuming any universal factor or determiner running through all possible human activities. It differs, on the other hand, not only from Spearman and Thorndike, but from Thomson as well,¹ in rejecting specific or unique factors. As to the universal factor, it may be regarded as doubtful whether Spearman himself has ever believed his general factor to be a constituent of every possible aptitude or ability. He seems rather to limit it to those activities which are "intellectual." For example, he remarks that "all branches of *intellectual* activity have in common one fundamental function." (Italics ours.) It is quite obvious, of course, that if there were numerous *non-intellectual* activities which did not involve this general intellectual factor, the latter would actually be nothing more than a wide-range group factor so far as a really comprehensive theory of aptitudes is concerned. In regard to the specific factors it may be said that if we assume even a single unique factor or determiner to be involved in each of the multitudinous activities that a human being is capable of performing, it would imply a prodigality of special organic mechanisms quite out of harmony with what we know of biological economy. Moreover, once group factors have been postulated, there seems no particular reason for assuming the existence of specific factors.

The implication of a strict group-factor theory for aptitude psychology may now be considered. Group factors, exactly

¹ In a private communication to the writer Thomson states that he is at present (1927) inclined to postulate group factors without either general or specific factors, but he seems to have published no definite statement of this position. In what is probably the most adequate account of his view (9) the example which he gives includes 13 general and 57 specific factors.

as a universal or general factor, permit of the possibility of finding tests which may correlate with aptitudes. But group factors possess the added theoretical advantage of the possibility that a test may correlate with one aptitude while not correlating at all with another. This is of great importance. Thus, group factors make it possible that independent tests *a* and *b* (Table 31) may correlate respectively with aptitudes *A* and *B* and yet at the same time neither test correlate with the other aptitude. This situation would result if test *a* were dependent upon some of the same group factors as aptitude *A* but not upon any found in aptitude *B*, while test *b* were dependent upon some of the same group factors as aptitude *B* but not upon any found in aptitude *A*; and lastly if test *a* contained none of the factors found in test *b*. Such a situation is shown in Table 31. Under such conditions, or under conditions approaching these, it clearly would be possible to tell within limits whether a given subject would have greater ability in aptitude *A* or aptitude *B* merely from knowing the scores in tests *a* and *b*. In other words, the existence of group

TABLE 31

ILLUSTRATING HOW GROUP FACTORS MAKE POSSIBLE THE DIFFERENTIATION OF APTITUDES BY MEANS OF TESTS

An inspection of the determiners constituting the four activities shows that with a given individual the score of test *a* might be high and that of test *b* might be low, since the two are not correlated. But because of the high correlation between each test and its aptitude, such test scores would indicate strongly that *A* would be high and *B* would be low. Thus the two aptitudes would be differentiated by means of the tests.

ACTIVITY	FACTORS OR DETERMINERS							
	I	II	III	IV	V	VI	VII	VIII
Aptitude <i>A</i>	×	×	×					
Test <i>a</i>	×	×						
Aptitude <i>B</i>					×	×	×	
Test <i>b</i>					×	×		

factors would permit the possibility of *differentiating* the potential aptitudes of an individual by means of tests. This, as we have already seen, a strict two-factor theory does not do.

It follows from the foregoing that a decisive indication of the existence of group factors would be furnished if it could be demonstrated that differences between the amounts of certain aptitudes possessed by an individual could be determined on the basis of ordinary test scores. This we have already seen (pages 42 ff.) appears to have been accomplished, though the matter needs a more rigorous investigation. It is doubly significant that this differentiation took place within the relatively restricted field of ordinary academic aptitudes. It should be much easier to accomplish where the aptitudes to be differentiated were more distinct.

THE THEORY OF APTITUDE OR BEHAVIOR LEVELS

Assuming the existence of complex group factors as determining particular aptitudes, we at once meet the question as to whether these group factors so combine as to produce certain distinct groups or levels of aptitudes. The aptitudes within a level would, it is assumed, correlate highly with other aptitudes in the same level but relatively low with aptitudes in other levels. Thorndike (88) has put forward such a theory, which assumes three kinds of intelligence: mechanical, social, and abstract. More recently Toops (94) has suggested the addition of a fourth, which he calls "clerical intelligence." Thorndike states his view of the relations existing among aptitudes as follows:

A perfect description and measurement of intelligence would involve testing the man's ability to think in all possible lines. . . . For ordinary practical purposes, however, it suffices to examine for three "intelligences" which we may call mechanical intelligence, social intelligence, and abstract intelligence. By mechanical intelligence is meant the ability to learn to understand

and manage things and mechanisms such as a knife, gun, mowing machine, automobile, boat, lathe, piece of land, river, or storm. By social intelligence is meant the ability to understand and manage men and women, boys and girls, to act wisely in human relations. By abstract intelligence is meant the ability to understand and manage ideas and symbols such as words, numbers, chemical or physical formulæ, legal decisions, scientific principles, and the like. . . .

Within any of these intelligences a man displays relatively great consistency. The man who learns carpentering quickly and well could commonly do nearly as well as a mason, sailor, plumber, millwright, or auto-repair man. The man who succeeds as a politician would commonly have done well as a salesman, hotel clerk, confidence man, or, if provided with certain accessory traits, as a parish priest or school principal. The boy who cannot learn algebra, history, and science will probably be unable to learn law, engineering, philosophy, and theology. Between one and another of these three is relatively great disparity. The best mechanic in a factory may fail as a foreman from lack of social intelligence. The whole world may revere the abstract intelligence of a philosopher whose mechanical intelligence it would not employ at \$3.00 per day.¹

The above statement of Thorndike's was unaccompanied by evidence and was probably intended merely as a shrewd conjecture or hypothesis. Recently, Freyd (21) has reported experimental results bearing on two of the supposed intelligences of Thorndike — the social and the mechanical. He made a rather intensive study of 30 men enrolled in a course for life-insurance salesmen. These were taken as a typical group who might be expected to possess more than average social intelligence. By way of contrast he carried out a similar study on 30 senior men enrolled in a technical, rather mechanical, course designed to fit men to be executives in industry. In addition, several other groups were examined for purposes of corroboration. Various psychological tests were given, together with a series of five rather elaborate

¹ Reproduced by permission of *Harper's Magazine*.

questionnaires. In many respects the two groups were found not to differ materially. There was, for example, no material difference between the two groups on a modified form of the Army Alpha intelligence test. But a number of characteristic differences were found in other directions. It was found that, on the whole, the sales group was able to disguise handwriting more readily than the industrial group. The salesmen were somewhat less positive of their correctness in a memory test, they were less self-conscious and less able to inhibit a tendency to write fast on a slow-writing test. On a written form of the Kent-Rosanoff Association Test it was found that the sales group had a distinctly greater tendency to commonality or conventionality of response. When asked to indicate their likings for a variety of vocations, the sales group with considerable uniformity showed a preference for those vocations in which they would be concerned mainly with people, such as acting, preaching, and especially salesmanship. The industrial group with equal consistency preferred dealings with inanimate materials. They were more attracted by the vocation of astronomer, locomotive engineer, tool maker, watchmaker, etc. Some curious differences showed up in various miscellaneous directions. The sales group tended to like fat men, the industrial group to dislike them. The sales group preferred the satirical periodical *Life*, while the industrial group preferred the serious, liberal *New Republic*. Rather reliable character ratings by the men themselves, and parallel ratings by other persons given as references, indicated that the salesmen tended to be more excitable, more self-confident, and more open-hearted, and that they made friends more readily. There were also tendencies, though somewhat less pronounced, for the salesman to be more wide-awake, more good-natured, more adaptable, more talkative, more neat, and less self-conscious than the industrial group. The sales group appeared to be distinctly more

credulous regarding popular pseudo-science, such as astrology and phrenology, and to accept uncritically popular opinion or slogans such as "All men are created equal" or "Unstinted service is always rewarded." The sales group wrote more social letters, learned to dance at an earlier age, and included a much higher percentage of smokers.

Taken as a whole, Freyd's results present a picture quite in harmony with general belief and expectation. They show pretty clearly that persons with certain traits tend to choose the one occupation and that persons with opposite traits tend to choose the other. So far as these traits go, the two groups are opposite or extreme. But since choice is not necessarily a guarantee of success in the vocation chosen, the exact significance of the various differentiating items as to vocational aptitude or success is not clear. In general, Thorndike's position regarding the three intelligences has not found much support. There is little evidence for types here except as extremes of distributions. Probably most persons would fall about midway between the extremes represented by the above two groups of subjects.

INTELLECTUAL AND MOTOR ABILITIES CONCEIVED AS LEVELS

Meanwhile there are other possibilities. As experimental results have accumulated there has gradually emerged a fairly marked distinction between intellectual abilities and intelligence tests on the one hand and motor abilities and motor tests on the other. This distinction is so marked that it suggests the two abilities as possible levels in Thorndike's sense. For these two classes of abilities to conform to Thorndike's criteria of separate intelligences, three conditions should be found: (1) There should be a high correlation among various tests of intelligence, (2) there should be little or no correlation between intelligence tests and motor tests, (3) there should be a high correlation among various motor

tests. The results from a rather large amount of careful experimentation show that two of the three conditions are approximately fulfilled. The third condition, however, pretty definitely is not.

1. It has long been known that intelligence tests show a marked tendency to correlate highly with one another. As a rather extreme example of this may be cited the correlations among the eight tests comprising Army Alpha. These results conform to the first condition.

TABLE 32

SHOWING THE CORRELATIONS AMONG THE VARIOUS TESTS IN THE ARMY ALPHA INTELLIGENCE TEST BATTERY. THESE COEFFICIENTS ARE BASED ON THE TEST SCORES FROM ABOUT A THOUSAND RECRUITS. (*Memoirs National Academy of Sciences*, 1921, Vol. XV)

No.	DESCRIPTION OF TEST	1	2	3	4	5	6	7	8
1	Directions	1.00	.73	.59	.71	.69	.68	.67	.66
2	Arithmetic	.73	1.00	.75	.79	.76	.77	.74	.74
3	Practical Judgment .	.59	.75	1.00	.81	.75	.61	.67	.78
4	Synonym-Antonym .	.71	.79	.81	1.00	.83	.68	.73	.86
5	Disarranged Sentences	.69	.76	.75	.83	1.00	.67	.78	.82
6	Number Series . . .	.68	.77	.61	.68	.67	1.00	.70	.69
7	Analogies	.67	.74	.67	.73	.78	.70	1.00	.67
8	Information	.66	.74	.78	.86	.82	.69	.67	1.00

2. The various forms of intellectual behavior show an equally marked tendency to correlate low with characteristic motor behavior. As illustration of this tendency may be cited the data of Table 33, which have been compiled from a variety of sources. Thus condition (2) is satisfied.

3. Motor tests, as a general thing, do *not* correlate highly with each other. This is contrary to condition (3) laid down above as demanded by a theory of levels. The fact seems first to have been noted by Perrin. He gave three complex motor tests and fourteen simple motor tests to 51 students at

TABLE 33

SHOWING THE CORRELATIONS BETWEEN VARIOUS FORMS OF INTELLECTUAL
AND MOTOR ACTIVITY

INTELLECTUAL ACTIVITY	MOTOR ACTIVITY	CORRELATION	AUTHOR
Army Alpha	Bogardus apparatus	+ .03	Perrin
" "	Card sorting	+ .02	"
" "	Coördinate tracing	+ .10	"
University marks	Bogardus apparatus	- .17	"
" "	Card sorting	+ .01	"
" "	Coördinate tracing	+ .21	"
Intelligence test	Test of physical efficiency (jumping)	+ .11	Engelhardt
University marks	Test of physical efficiency (jumping)	+ .29	"
Army Alpha	Motor ability as rated by three gymnastic teachers	+ .02	Garfiel
Intelligence test	Stenquist Mechanical As- sembly	+ .18	Toops
Army Alpha	Stenquist Mechanical As- sembly	+ .14	O'Rorke

the University of Texas. His published correlation coefficients are shown in Table 34. By way of explanation it may be stated that the Bogardus test is a rather complex apparatus designed to reproduce some of the essential conditions where a workman tends an automatic machine. The co-ordination test measures the ability of a person to trace accurately at a given speed the outline of a triangle and a square, one with each hand and simultaneously. The inhibition test was designed to test the power of inhibiting the winking of the eyelid when piano hammers struck a plate of glass upon which the face was resting.

Muscio (59) continued the investigation with various groups of young people from London, using relatively simple motor tests. Where the two hands were correlated for the same activity the coefficients were fairly large, ranging around

TABLE 34

SHOWING CORRELATIONS AMONG VARIOUS MOTOR TESTS (After Perrin)

"COMPLEX" MOTOR TESTS	BOGARDUS r	CARD SORTING r	COÖRDINA- TION r
1. Bogardus			— .10
2. Card sorting	— .06		+ .36
"SIMPLE" MOTOR TESTS			
1. Reaction time	+ .16	+ .21	— .06
2. Inhibition of winking reflex .	+ .02	— .11	+ .02
3. Motor memory	+ .02	+ .22	+ .17
4. Weight discrimination . .	— .07	— .03	— .01
5. Aiming	+ .24	— .09	+ .12
6. Aiming (blindfolded) . . .	+ .29	+ .05	+ .04
7. Balancing	+ .38	— .06	— .11
8. Rhythm (pendulum) . . .	+ .17	+ .20	+ .29
9. Rhythm (kymograph) . .	— .02	+ .12	+ .12
10. Tapping	+ .22	+ .02	+ .01
11. Steadiness	+ .09	— .04	+ .01
12. Tracing	— .11	— .01	+ .09
13. Dynamometer	— .22	+ .20	+ .39
14. Vital capacity	— .14	+ .09	+ .22

+ .40. Otherwise his coefficients were about the same as Perrin's. Muscio concludes, "... an individual's performance in one such activity is not in general the slightest indication of what his performance in another such activity will be."

As a final piece of evidence may be cited the careful experimental study of Ella Barton (2) on the correlations among motor abilities. She tested with scrupulous care 75 university women ranging in age from 18 to 23 years. Each subject was tested twice on each of 14 test activities, mostly of a simple motor type. The total of the two trials was taken as the test score. The correlations of each test with all the

TABLE 35

SHOWING THE CORRELATIONS AMONG VARIOUS TESTS, MOSTLY OF SIMPLE MOTOR SPEED PROCESSES EXECUTED BY THE HAND (Barton)

NAME OF TEST	No. OF TEST	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Easy directions test (W + W) . . .	1														
Speed of oral counting	2		-.01	+.11	.00	-.18	+.23	+.21	+.22	+.23	+.04	+.30	+.04	+.20	+.31
Speed of writing . . .	3			-.11	+.10	+.09	+.20	+.19	-.01	+.23	-.05	+.19	+.02	-.04	+.24
Speed of canceling A's	4				+.58*	-.01	+.08	+.13	+.37	+.06	-.06	+.07	+.16	-.13	-.19
Card sorting . . .	5					+.07	-.19	+.28	+.20	+.07	+.08	-.01	-.03	-.08	+.03
Tapping with tele- graph key . . .	6						-.02	+.15	+.19	+.08	-.05	-.21	+.11	+.05	-.18
Tapping with pencil .	7							+.43*	+.18	+.02	+.09	+.46*	+.43	-.10	+.16
Putting pegs in holes	8								-.23	+.20	-.22	+.11	+.47	+.13	+.28
Taking pegs from holes	9									+.35*	+.38*	+.29	+.20	-.28	+.05
Speed of buttoning .	10										-.20	+.05	+.08	+.09	+.10
Speed of unbuttoning	11											+.50*	+.04	-.23	-.01
Three-hole test — successes	12												+.11	-.11	+.25
Three-hole test — misses	13													-.42*	+.11
Whipple steadiness test	14														+.01

others were then computed. The Hull correlation machine was employed, together with the system of checks described on pages 427 to 439, so that the absolute accuracy of the computations is assured. The coefficients were then all corrected for attenuation, though the corrections were usually slight. The reliability coefficients of the combination of two repetitions of the various tests in most cases was around .90. The resulting corrected correlations are shown in Table 35. The signs of these r 's have been freed from any eccentricities due to peculiarities of scoring, all coefficients having been given the sign that they would have received had a good or desirable score always been a large number and a poor score a small number.

In general, Barton's results harmonize with those of the other workers in motor tests, though her coefficients in certain cases attain considerable size. Several of the larger ones have been marked with an asterisk. The correlation of $+.58$ between speed of writing and the cancellation of A's is noteworthy. Several of the other large r 's are obviously the result of special choice of tests. Thus tapping with a telegraph key and tapping with a pencil correlate $+.43$; placing pegs in holes correlates with taking them out to the extent of $+.35$. When one considers the obvious elements of similarity between these pairs of activities, it is really surprising that these particular correlations are not higher. On the other hand, 27 of the 91 r 's are negative. For the most part these negative coefficients are small and probably do not signify true negative relations, but rather represent correlations approximating zero.

Summarizing the evidence as to the relations among motor activities, we may say that as a rule they correlate very low with each other; but they occasionally show rather high correlations. This is likely to be the case when the apparent similarity between the two activities is very marked. These

facts seem effectually to preclude belief in a two-level theory of aptitudes made up of complex intellectual and simple motor abilities.

COMPLEXITY OF BEHAVIOR AS A FACTOR

The results of Perrin, Muscio, and Barton show pretty clearly that simple motor tests yield remarkably slight correlation with each other when compared with the correlations usually obtained from various tests of "intelligent" behavior. In significant contrast stand the results of McFarlane (54). She tested a number of groups of boys and girls, chiefly from the London schools. She used as tests (1) the assembly of a small wooden wheelbarrow and a small wooden cradle, (2) the assembly of two garments the parts of which could be snapped together with metal fasteners, (3) a 3-inch painted cube-assembly test, (4) the Healy puzzle box, and (5) the McDougall "plunger" test. The apparatus of this consists of 24 metal sockets arranged in a circle, and a plunger or stylus slightly smaller than the sockets. The plunger is inserted into one socket after the other as rapidly as possible. The results from one of her characteristic groups of subjects (40 boys, age 14) are given in Table 36. These coefficients differ from those of Perrin and Muscio in showing, upon the

TABLE 36

SHOWING THE CORRELATIONS AMONG VARIOUS COMPLEX MOTOR TESTS
(After McFarlane, '25)

	GAR- MENT ASSEM- BLY	CUBE ASSEM- BLY	HEALY PUZZLE BOX	PLUNGER TEST	EXAMI- NATION MARKS	INTELLI- GENCE TEST
Wood assembly . .	+.32	+.32	+.32	+.02	-.11	-.26
Garment assembly .		+.46	+.44	+.19	+.11	-.04
Cube assembly . .			+.44	-.07	+.04	+.14
Healy puzzle box .				+.11	+.07	+.26
Examination marks .						+.75

whole, distinct correlations among the motor tests. Her other groups of subjects show very similar values. It is interesting to note that at the same time the motor tests correlate markedly less with the two intelligence criteria, though the latter correlate with each other to the extent of $+.75$. McFarlane believes that her correlation coefficients are so much larger than those of Perrin and Muscio because her tests (with the exception of the plunger test) involve much more complex behavior.

Upon the whole the evidence so far available does not reveal any very clearly marked levels of aptitudes such as Thorndike has suggested. Indeed, it is doubtful whether any aptitude levels in a strict sense, having high correlations among the activities on a given level and a low or zero correlation with aptitudes on all other levels, ever will be found. It is almost certain, however, that for practical purposes of aptitude prediction a rather large number of groups or constellations of aptitudes ultimately will be isolated. It should be observed that groups would result from the action of chance alone. If pebbles should be tossed into the air, they would fall in such a way that they would be divisible into groups or constellations very much as has been done in the case of the distribution of the stars in the sky. Yet it must be remembered that aptitudes are not in a two-dimensional space, but possibly in a multi-dimensional one, and that the situation is consequently not nearly so simple as the above expression might imply.

While experimental results have failed to show any convincing evidence for distinct levels of abilities, it has revealed what may turn out to be a significant tendency or law. This is that *the highly complex intellectual activities correlate highly with each other, the less complex correlate with each other to an intermediate degree, and the relatively simple motor activities correlate with each other only slightly.*

The hypothesis that diminishing correlations tend to result from diminishing complexity of activity finds a rather ready explanation in the theory of group determiners. No one knows how many aptitude determiners there are. For purposes of definiteness, let us assume 1300 as the total number. It is quite clear that two complex intellectual activities, each involving several hundred determiners, would *by mere chance alone* have in general a large number of determiners in common and thus a high correlation. On the other hand, the likelihood would be small that two simple activities, each involving only a few score of determiners, would contain by chance many determiners in common. Fortunately it is possible to tell what correlations would result in the long run from various assumed conditions. Computation shows that an activity complex enough to involve a random sampling of any 800 of the total 1300 determiners would, on the average, correlate $+.61$ with others of the same degree of complexity. Activities which would involve a random sampling of any 400 of these determiners would on the average, by mere chance, overlap others of the same degree of complexity sufficiently to correlate $+.31$.

TABLE 37

SHOWING THE DEGREES OF CORRELATION WHICH WOULD RESULT ON THE AVERAGE FROM THE CHANCE OVERLAPPIINGS OF THE DETERMINERS OF ACTIVITIES HAVING VARYING DEGREES OF COMPLEXITY

It has been assumed here, for purposes of simplicity in computation, that all determiners are equally weighted.

TYPES OF ACTIVITY	DEGREE OF COMPLEXITY ASSUMED. NUMBER OF DETERMINERS OUT OF HYPOTHETICAL TOTAL OF 1300	DEGREE OF CORRELATION RESULTING FROM CHANCE OVERLAPPIINGS OF DETERMINERS
Intelligent	800	$+.61$
Complex motor . .	400	$+.31$
Simple motor . . .	50	$+.04$

Activities which are so simple as to involve only 50 determiners would overlap sufficiently by chance to correlate on the average $+.04$, or almost not at all. The correlations resulting from these overlappings are of about the size found among "intelligence" tests, complex motor tests, and simple motor tests, respectively. Thus, some of the peculiarities found in test results can be accounted for very largely as resulting from the mere complexity of the tests themselves.

It must be understood that the above holds only for the average of many correlations. Undoubtedly there are certain wide-range determiners largely common to most intelligence tests, corresponding to the large verbal and symbolic elements involved. Among most group intelligence tests such as Army Alpha, there is the large additional element of pencil-and-paper reactions. There are probably similar group determiners of wide range running more or less through complex motor activities of the sort called for by various assembly tests. Such additional common group factors would, of course, increase the correlation among the aptitudes containing them, above the chance overlappings noted above. On the other hand, chance would frequently produce much less than the average amount of overlapping for a given complexity of determiners constitution. This would account for occasional very low correlations among fairly complex forms of behavior.

CHAPTER SEVEN

FUNDAMENTAL RELATIONS AMONG APTITUDES AND TESTS

IN the last chapter we examined various views as to the nature and constitution of aptitude and test behavior. It will be recalled that in that connection we were led by the evidence to the adoption of a strict group-factor theory. Since this view is made the basis of the exposition of the relations among aptitudes and tests contained in the present chapter, it will be convenient at this point to state its principles and assumptions somewhat more fully.

THEORY OF GROUP APTITUDE DETERMINERS

As already suggested, the basic concept of the theory is that of group factors, or determiners of efficiency. These determiners, or factors, unite in various combinations to produce the various aptitudes which an individual possesses. The same determiner in most cases will contribute to the success of numerous different activities. Some determiners, perhaps, may be constituents of no more than three or four activities, while others may contribute to the efficiency of several hundred. The former may be narrow-range group factors, approaching the Spearman specific factors; the latter would be wide-range group factors, approaching the Spearman general factor.

Each determiner within a given person is possessed of a certain specific degree of potency, or physiological efficiency. This is assumed to remain constant, no matter in what activity the determiner may enter or with what other determiners it may be associated in particular aptitudes (see Table 39). The degree of the physiological efficiency of each determiner possessed by an individual is totally unre-

lated to that of any other determiner; i.e., *all determiners correlate exactly zero* with the others.¹

But while each determiner possesses a given physiological efficiency within a given individual (at any particular time), the relative importance of that determiner in contributing to the efficiency of the different activities into which it enters may vary widely. For example, if the pitch discrimination mechanism of the inner ear should turn out to be an elementary aptitude determiner, it might easily be found (1) to be of great importance in tuning pianos, (2) to be of some importance in playing the bass drum, and (3) to be of no importance whatever in professional tea tasting. In other words, a particular ultimate physiological determining mechanism will be weighted more or less differently in every aptitude in which it plays a part. Similarly, the different determiners of any particular aptitude will ordinarily be weighted very differently from one another. This makes our theory one of *weighted* group factors or determiners.

In the following pages these weighted aptitude determiners will be treated as if their mode of combination in the formation of individual aptitudes were that of the simple addition of their several weighted values. Thus, the aptitude of a person for a particular activity is supposed to be the simple summation of the products of his various relevant determiner potencies, each multiplied by its characteristic weighting or importance in the aptitude in question (see Table 39). There is no special *a priori* reason for expecting this particular

¹ The reasons for this assumption may be indicated briefly. Assume that two supposedly distinct determiners should be found to correlate to a certain extent with each other. The question would naturally arise as to the reason for this tendency to vary together. The explanation must be found in some factor common to the two supposed determiners. This common determining source must naturally be a still more elementary mechanism. This means that the supposed determiners are really aggregates, themselves in need of a correlation explanation, and obviously are not suitable entities for explaining other correlations.

mode of combination to obtain, any more than that of numerous others. It may very well be that a number of different principles of combination are operating at once. It is a fact, however, that the various observed relations among aptitudes and tests may be accounted for to a remarkable degree by merely assuming numerous group determiners, of varying range and weighting, which combine by summation to determine the efficiency of the most diverse forms of activity.

It must be admitted that as yet almost nothing is known as to the identity of the ultimate aptitude determiners. We have suggested above that the mechanism of the inner ear which mediates the discrimination of pitch may be such a determiner. The reader may at least take this as a concrete example to illustrate our conception of one type of determiner. Probably there are, however, numerous different types and orders of determiners. This lack of concrete information regarding the ultimate elements of our system is suggestive of the theories in the physical sciences employing atoms, molecules, and ether as explanatory concepts. Materials behave *as if* they were composed of, or dependent upon, these more or less hypothetical entities. And we find as a matter of fact that reasoning based on the assumptions of atoms, ether, etc., leads to further observations and illuminating discoveries and thus is justified. In this sense the present theory of aptitudes is also a kind of atomism. The various observed relations among aptitudes and tests appear to be such as *would* be produced by independently weighted group determiners.

THEORETICAL DETERMINER ANALYSIS OF A CONCRETE TEST-APTITUDE RELATION

We shall now apply the above method of analysis to a concrete case. Dr. A. T. Weaver tested the pitch discrimi-

nation of a man who was taking special work in public speaking. The man was noted for having a remarkably monotonous speaking voice. The test revealed him to be almost entirely lacking in ability to discriminate pitch. This suggested the possibility to Dr. Weaver that the ability to discriminate pitch might be an important component of the power of vocal expression. He accordingly measured in a most painstaking manner the capacity of a large number of young men to read expressively a difficult passage. He then tested the same individuals for pitch discrimination by the Seashore phonograph method. The two capacities were found to correlate to the rather surprising extent of $+.48$.

Now, how may we conceive these two processes to be organized so as to correlate in this way? In the first place there is no reason to believe that an aptitude activity such as interpretative reading is constituted in essentials differently from a test activity such as pitch discrimination, except that probably aptitude activities are, in general, somewhat more complex. *By complexity is here meant merely the dependence upon a greater number of determiners.* Let us say, then, that vocal expression is dependent upon six determiners, while pitch discrimination is dependent upon three.¹ In order to secure a moderate amount of correlation, let us assume that two determiners are common to the two activities. We may represent the situation conveniently as in Table 38. Ten of the many possible determiners are represented by Roman numerals. Determiners V and VI are assumed to be common to both activities. Of these, V is of major importance in both activities, as is indicated by the relatively large weight of 4 in vocal expression and of 5 in pitch discrimination. This may plausibly be considered as

¹ It is not intended by this example to imply that such a small number of determiners is involved. The number is probably really scores or even hundreds. The small numbers are taken for convenience in exposition.

TABLE 38

SHOWING A HYPOTHETICAL ARRANGEMENT OF DETERMINERS FOR VOCAL EXPRESSION AND PITCH DISCRIMINATION WITH WEIGHTINGS SUCH AS WOULD BRING ABOUT A CORRELATION IN HARMONY WITH EMPIRICAL OBSERVATION

ACTIVITY	DETERMINERS									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Weights, vocal expression (0) . .	7	3	1	1	4	2				
Weights, pitch discrimination (1) .					5	1	1			

representing the pitch discrimination mechanism of the ear, already mentioned. Determiner No. VI may be considered as an additional common determiner as yet unidentified and unsuspected because of our meager knowledge of the essential mechanics of the two processes. Determiner I, the most important component of vocal expression as shown by its relatively great weight of 7, is not represented in the test at all. It may be thought of, perhaps, as some function of vocal coördination. Incidentally it appears that a relatively small number of determiners of vocal expression need be sampled by the test and that even these determiners need not be similarly weighted in the two activities in order to produce approximately the observed correlation of .48.

HOW COMMON APTITUDE DETERMINERS PRODUCE CORRELATIONS

Let us now observe in some detail just how the aptitude determiners common to the test and the aptitude as represented in Table 38 produce a correlation. It will be remembered that the efficiency of each determiner within a particular individual organism is regarded as a constant quantity,

TABLE 39

SHOWING THE DERIVATION OF CORRELATED SCORES FROM DICE THROWS
AS FOR HYPOTHETICAL SUBJECT NO. 1, TABLE 40

	DETERMINERS							FINAL SCORE
	I	II	III	IV	V	VI	VII	
Weights, vocal expression (0)	7	3	1	1	4	2		
Weights, pitch discrimina- tion (1)					5	1	1	
Determiner efficiencies of a hypothetical subject No. 1 of Table 40 as fixed by dice throws								
Weighted values of deter- miners, vocal expression .	7	12	6	6	12	2		= 45
Weighted values of deter- miners, pitch discrimina- tion					15	1	1	= 17

no matter in what aptitude it may be found. Since the various determiners are strictly uncorrelated, we may fix once and for all their respective efficiencies for hypothetical single individuals, by means of dice throws. Thus we may lay open for observation the mechanism of the correlation. Suppose, then, we have seven dice numbered from I to VII corresponding to the respective determiners represented in Table 39. Upon the first throw, the dice showed the faces represented in the middle line of Table 39. The numbers indicated by these dice faces may accordingly be taken as the strengths or efficiencies of the corresponding determiners of a hypothetical subject. The aptitude weights of the various determiners for vocal expression are shown in the first line and those for pitch discrimination in the second. The

product of each determiner efficiency (dice face) multiplied by its aptitude weight in vocal expression is given in the fourth line. Thus:

$$7 \times 1 = 7$$

$$3 \times 4 = 12$$

$$1 \times 6 = 6, \text{ and so on.}$$

The corresponding products for pitch discrimination are given in the last line. These weighted values, when added, yield the efficiencies of the individual in the aptitude and the test respectively:

$$7 + 12 + 6 + 6 + 12 + 2 = 45 \text{ (score in vocal expression)}$$

$$15 + 1 + 1 = 17 \text{ (score in pitch discrimination)}$$

Twenty-five dice throws, similar to and including the one just worked out in detail, are shown in Table 40. In the two columns at the right are shown the corresponding resulting scores beginning with the 45 and 17 already worked out above. An inspection of these two columns of scores shows that, while there are exceptions, as a general thing a large number in the first column corresponds to a relatively large number in the second. This is because of the fact that if determiners V and VI chance to be rather large in the aggregate, then, other things equal, both resulting totals will be large in about the same proportion, since both are alike dependent on these particular determiners. The reason that the correlation is not perfect is due largely to the fact that the non-common determiners I, II, III, IV on the one hand, and VII on the other, varying quite independently, may largely mask the agreement actually existing in the common determiners. In individual cases this may even result in values quite reversing the general tendency. Thus, subject No. 8 (Table 40) scores 63 in vocal expression and only

TABLE 40

SHOWING THE MECHANICAL SYNTHESIS OF A CORRELATION OF $+.465$
WHERE THE DETERMINERS ARE WEIGHTED AS ASSUMED ABOVE
(PAGE 223) FOR VOCAL EXPRESSION AND PITCH DISCRIMINATION

The values of the determiners for the 25 hypothetical subjects were found by dice throws as described in the text.

HYPOTHETICAL SUBJECT'S NUMBER	DETERMINER VALUES FOUND BY DICE THROWS							SCORES IN THE CORRELATED VARIABLES	
	I	II	III	IV	V	VI	VII	"Vocal Ex- pression" (0)	"Pitch Dis- crimination" (1)
1	1	4	6	6	3	1	1	45	17
2	1	4	2	3	5	1	4	46	30
3	5	3	6	3	2	1	1	63	12
4	2	1	4	2	2	4	2	39	16
5	6	5	1	2	6	4	5	92	39
6	1	3	2	2	2	1	3	30	14
7	4	4	2	4	5	5	6	76	36
8	6	3	5	1	1	1	4	63	10
9	2	1	6	3	3	2	1	42	18
10	4	2	4	2	6	2	1	68	33
11	6	3	2	1	3	1	3	68	19
12	4	4	1	6	4	6	1	75	27
13	6	3	2	1	4	2	6	74	28
14	4	6	4	3	5	1	4	75	30
15	1	1	5	1	3	4	4	36	23
16	1	2	3	6	5	2	1	46	28
17	3	1	3	2	1	6	1	45	12
18	6	5	4	2	2	3	6	77	19
19	1	2	1	6	4	3	5	42	28
20	5	4	5	5	2	6	2	77	18
21	4	3	2	4	6	4	5	75	39
22	3	4	5	5	1	6	2	59	13
23	6	5	3	2	3	1	5	76	21
24	4	3	3	5	5	4	5	73	34
25	1	4	1	4	2	5	6	42	21

10 in pitch discrimination. This is because the common determiners happen to run small. The non-common determiner, No. VII, while of fair size, has a very small

weight. This makes the test score very small. The non-common determiners of the aptitude, particularly I, II, and III, are, however, very large. This gives a large aptitude score in spite of the small size of the common determiners. Chance combinations of the non-common determiners extremely favorable to the correlation may also occur. Such cases are seen in the scores of subjects 5 and 6, in one both scores being large and in the other both small.

The correlation of the two columns of scores turns out upon computation to be $+.465$, which is very similar to that of the empirically observed correlation between vocal expression and pitch discrimination.

ELEMENT OF CHANCE IN DETERMINING SIZE OF CORRELATION COEFFICIENTS

The large element played by chance in correlations, as shown above, makes it at once clear that the scores of any small group of individuals are inadequate to reveal the true relations existing between the variables. This is shown quite impressively by the great variety of correlations which may be obtained in successive series of 25 dice throws. The writer had 5000 dice throws made, and correlations were computed of the successive series of 25 in the manner already described. The highest r obtained was $+.77$ and consequently in error by 30 points, the smallest $-.09$ with an error of 56 points! The actual distribution of the 200 obtained coefficients is shown in Figure 32. It is sufficiently evident from this enormous range of values obtained from an *identical correlational situation* that coefficients are subject to large errors due to the accidental nature of the samples secured. If, however, the size of the groups is quadrupled, the variability of the obtained coefficients is reduced by approximately half, as shown by Figure 33. Finally, when all 5000 cases are taken at once, the r obtained is $+.4712$,

which deviates by only .0022 from the true theoretical value that would result from an infinite number of cases.

The moral of the preceding paragraph is that correlation coefficients are dependent to a considerable extent upon the accidental characteristics of the samples of data from which they are computed. It is clear that this accidental element grows less and less with the increase in the size of the samples taken. Accordingly the samples taken should always be as large as circumstances permit. The accidental element is also dependent in part upon the size of r . The size of this total chance factor is commonly expressed by what is called the *probable error*. This important statistical value is expressed by the formula,

$$\text{P.E.}_r = .6745 \frac{1 - r^2}{\sqrt{N}} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where r is the correlation coefficient and N is the number of cases.

The significance of the probable error of the correlation coefficient will become clearer if we first substitute in the formula the value of r (.4734) and of N (25). Substituting and solving, we have,

$$\begin{aligned} \text{P.E.}_r &= .6745 \frac{1 - .4734^2}{\sqrt{25}} \\ &= .1045 \end{aligned}$$

This means that, according to the principle of chance, the correlations obtained from such a set-up as was assumed in the case of the dice-throwing experiment, we should expect about half to deviate from the true value (+.4734) by less than .1045 and half to deviate by more than .1045. By referring to Figure 32 it may be seen that the value yielded by the formula agrees closely with the facts. It was found by actual count that 90 coefficients, or slightly less than half of the 200 in this series, were in error by more than .1045,

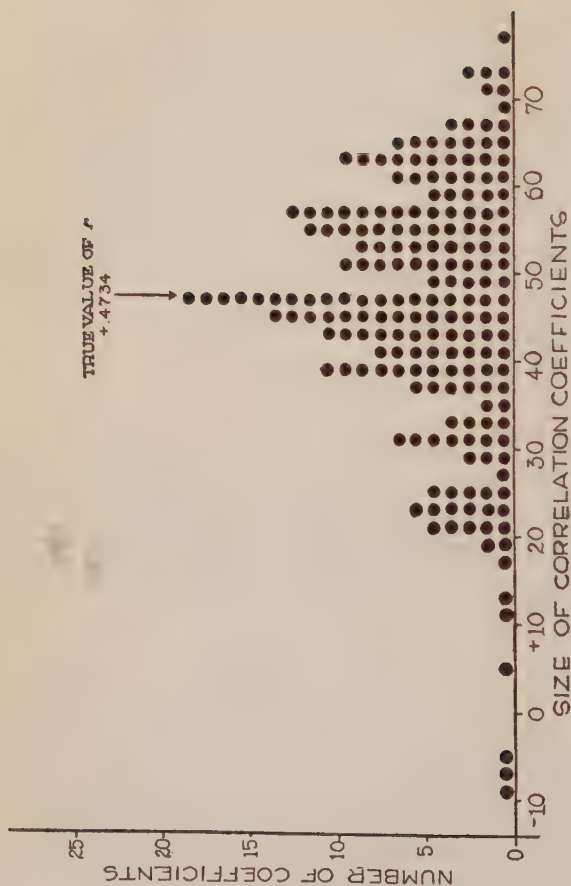


FIG. 32. This figure illustrates the unreliability of the correlation coefficient as an index of relationship when N is small. It shows the distribution of 200 correlation coefficients, each derived from 25 dice throws. All throws were made under exactly similar conditions and (except from errors due to the small N 's) should have given the same r , +.4734. Each dot represents a correlation coefficient. (See Figure 33.)

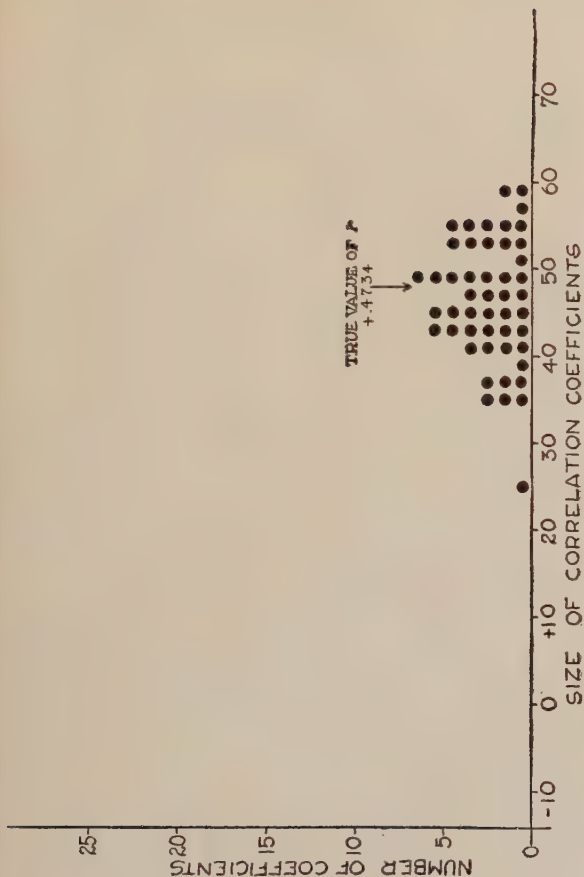


FIG. 33. This figure shows the distribution of 50 correlation coefficients, each derived from 100 dice throws. These coefficients are based on exactly the same 5000 throws as those of Figure 32, only taken in groups four times as large. Notice how much more closely they cluster around the true value. The median error is reduced to about half that shown in Figure 32, which is in accordance with theory. When all 5000 cases were taken at once, there was practically no error at all, the obtained r being $+.4712$.

while 110 were in error by less. It is important also to observe that 33 coefficients or $16\frac{1}{2}$ per cent of the whole were in error by more than twice the P.E., that six coefficients or 3 per cent were in error by more than three times the P.E., and that three coefficients or $1\frac{1}{2}$ per cent were actually in error by more than .417, or four times the P.E. It is also noteworthy that while these errors tend clearly to conform to the "normal" form of distribution, this distribution is not symmetrical. It tails off toward zero, as may easily be seen by a glance at Figure 32. This means that obtained r 's are a little more likely to be too small than too large.

When the probable error formula is applied to the group of 50 coefficients derived from 100 cases each (Fig. 33), substitution in the formula shows that the theoretical P.E., is .0522, or exactly half that where N is 25. Here again the facts agree closely with the theoretical value. With this group of coefficients it was found that 26 coefficients were in error by more than the P.E. while 24 were in error by less, a difference of only one from the theoretical expectation. Moreover, nine coefficients or 18 per cent were in error by twice the P.E., while two coefficients or 4 per cent were in error by three times the P.E. Here again the distribution tends to be "normal" and the tendency to tail off toward zero is present, though much diminished. All of the above results are thoroughly typical of the size and general distribution of the chance errors to which correlation coefficients are subject on the basis of the size of the samples of data from which they are computed.

Several times in the preceding discussion we have spoken of the true correlation which should be obtained from the combination of determiners if an infinite number of subjects (or dice throws) should be used. This is found by the following method, which is said to have been first derived by Poisson: *The sum of the products of pairs of common deter-*

miner weights divided by the square root of the product of the sum of the squares of the weights of one variable by the sum of the squares of the weights of the second variable. Indicating the operations and using values found in Table 39, we have:

$$\begin{aligned} r &= \frac{4 \times 5 + 2 \times 1}{\sqrt{(7^2 + 3^2 + 1^2 + 1^2 + 4^2 + 2^2)(5^2 + 1^2 + 1^2)}} \\ &= \frac{22}{\sqrt{2160}} \\ &= .4734 \end{aligned}$$

For purposes of theoretical analysis this formula is obviously superior to the actual throwing of dice. It will accordingly be utilized henceforth, wherever groups of determiners are set up for purposes of producing correlations.

THE NATURE OF ATTENUATION

Having before us the general nature of aptitudes and tests and the mechanism by which tendencies to correlation are presumably produced, we may turn to the consideration of certain theorems or fundamental principles bearing on the technique of aptitude testing. Several of these relate to the subject of *attenuation*.

If a test is given twice to the same group of subjects and the second set of scores is correlated with the first, the correlation is never perfect. It is often very far from perfect. The same is true of independent comparable measures of aptitude efficiency to an even more marked degree. Such correlations between independent measures of aptitude or test capacity are called "reliability coefficients." In practice these coefficients range from zero up to nearly perfection, or +1.00. The minimal reliability coefficient of measures usually considered necessary for use varies considerably with circumstances. Usually a test is not considered of much value if its reliability coefficient falls below

.50, and the same may also be said of aptitude measures as well. It is much more difficult to secure satisfactory reliability coefficients for aptitude criteria. For this reason we do not know so much about how they run. It is probable, however, that aptitude criteria as a class have much lower reliability coefficients than do tests.

The fact that in aptitude testing reliability coefficients are never perfect means that measurements of behavior functions are always to a certain extent defective. As a consequence of this ever present defectiveness or inaccuracy of measurement, all correlation coefficients obtained in aptitude work are somewhat reduced or attenuated below the figure which a true measure of the function would have yielded. This means that all raw or uncorrected coefficients obtained in testing work are an understatement of the actual relations of the function investigated. This shrinkage of the correlation coefficients is called *attenuation*.

It is probable that at least two kinds of attenuating forces are operating to diminish the coefficients in question: (A) inadequate sampling of the various activities constituting a compound aptitude or test, and (B) the presence of various irrelevant influences of accidental origin, such as fortuitous interruptions, etc. Very likely both types of attenuating influences are in most cases operating simultaneously. For purposes of analysis, however, it will be convenient to consider the two separately.

The inadequate sampling of the various component activities of a complex process (such as an omnibus test, say) would evidently be likely to bear somewhat more heavily on certain determiners on one occasion than on another. Such a situation may readily be represented as in Table 41. By the method used above (pages 230-231) the reliability coefficient in the assumed situation would be $+.68$. This is a rather mediocre reliability for a

TABLE 41

SHOWING THE RESULTS OF INADEQUATE SAMPLING AS PROBABLY INFLUENCING THE WEIGHTING OF DETERMINERS ON TWO REPETITIONS OF A TEST OR APTITUDE CRITERION, $r = +.68$

ACTIVITY	DETERMINERS							
	I	II	III	IV	V	VI	VII	VIII
First repetition . . .		2	1	3	4		7	
Second repetition . .		5	2	4	6		1	

test but is rather good for a criterion score, as such reliabilities run.

It may be remarked in passing that Table 41 illustrates an extremely important point not particularly related to attenuation. It is that two activities may be dependent upon exactly the same determiners and yet show a comparatively low correlation, solely because of the difference in the weighting or importance of the several determiners. It is probable that some such mechanism as that shown in Table 41 is responsible for the surprisingly low correlations often found to exist between processes which seem to superficial observation almost identical. Despite the identity of the determiners, the actual forecasting efficiency of one variable by the other in the above situation is less than 28 per cent (pages 268 ff.). There is evidently no need for assuming specific determiners (factors) to account for low correlations.

The influence of various irrelevant factors of accidental origin upon the correlation of successive or independent measures of the same activity is shown in Table 42. The irrelevant factors are all assumed to be uncorrelated, which must be the case if they are truly accidental. The correlation in this case is $+.67$, or very nearly that of the case just considered.

TABLE 42

SHOWING THE EFFECTS OF NUMEROUS SMALL IRRELEVANT FACTORS ON THE RELIABILITY COEFFICIENT $r = +.67$

ACTIVITY	DETERMINERS					IRRELEVANT FACTORS									
	I	II	III	IV	V	I	II	III	IV	V	VI	VII	VIII	IX	X
First measure		2	1		3	1		1		1		1		1	
Second measure		2	1		3		1		1		1		1		1

TABLE 43

SHOWING EFFECTS OF REPEATING BOTH TESTS SHOWN IN TABLE 42 ON THE RELIABILITY COEFFICIENT $r = +.80$

ACTIVITY	DETERMINERS				IRRELEVANT FACTORS																			
	I	II	III	Δ	I	II	III	Δ	I	II	III	Δ	I	II	III	Δ	I	II	III	Δ	I	II	III	Δ
First measure doubled		4	2		6	1																		
Second measure doubled		4	2																					

HOW MORE ADEQUATE MEASUREMENT REDUCES ATTENUATION

It is always assumed in practical testing work that if a test is repeated, the results of the two repetitions combined will give a more accurate measure than either alone. It may be shown that the combined results of two repetitions of a test will correlate with the combined results of two other equally good repetitions of a test to the extent of

$$\frac{2r}{1+r} \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (2)$$

where r is the correlation between any two single repetitions of the test. In case $r = +.67$, as in the example of the fallible measure just considered, the formula yields a value of $+.80$, which indicates a very considerable improvement in the accuracy of the measurement as the result of taking two repetitions of the test instead of one.

Let us now consider the inner mechanics of this improvement in the accuracy of measurement obtained by a repetition of the test, from the point of view of the determiners. If we combine two repetitions of a measure such as represented in Table 42 by addition, there is obtained an interesting result. The weights of the true determiners become twice as great as before, while the irrelevant factors remain of the same size but become twice as numerous. Two sets of such determiners obtained in this way are shown in Table 43. The correlation of these sets of determiners is found by direct computation to be $+.80$, exactly as given by the formula considered above.

At first it may appear paradoxical that the correlation should rise at the same time that the irrelevant factors have increased so greatly in number. The explanation lies in the fact that the true determiners have at the same time increased in relative weight or importance. The moral

would seem to be that *the more finely divided the irrelevant or attenuating factors are, the less damage they will do in aptitude testing.*

By a logical extension of the results obtained above, we may say that the greater the number of independent repetitions of a test which are combined into a single score, the more accurately the activity in question is measured. This is practically equivalent to saying that the longer the time spent in the test activity by the subject, the more accurately will his activity be measured. It often happens, however, that a measure cannot be repeated because of lack of time or because of some other circumstance connected with the nature of the experimental situation. In such cases the test results can sometimes be divided into halves, and each half then treated as if it had been a separate measure. This is especially useful in offering the possibility of determining the reliability of the test. It must be observed, however, that the correlation of the two halves of a measure does not yield the reliability coefficient of the measure, but rather of the *halves* of the measure. In such cases the reliability coefficient of the measure itself (the correlation the measure would show with another equally good) is yielded by the formula already considered :

$$\frac{2r}{1+r}$$

in which r is the correlation between the two halves of the original measure.

It occasionally happens that the reliability of a test or of an aptitude criterion is not satisfactory, and it may be desired to know how high its reliability will rise as a result of combining various numbers of repetitions. This is given by the Spearman-Brown formula of which the formula

just considered is a special case. The Spearman-Brown formula is :

$$\frac{rN}{1 + r(N - 1)} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where r is the reliability coefficient and n is the number of repetitions of the measure which are combined into a single score. In the case of the test situation represented in Table 42, where the reliability coefficient is $+.67$ and N is 5, this formula becomes :

$$\frac{.67 \times 5}{1 + .67(5 - 1)} = .91$$

This means, in plain words, that if five equally good repetitions of this measure were combined they would correlate with the combined scores of five other equally good repetitions of the measure to the extent of .91. This has been found by computation based on determiners to be equivalent to reducing the seven irrelevant factors of Table 42 to between one and two. By substitution in the formula in a manner similar to that for five repetitions, it may be seen that the same measure with 25 repetitions combined would yield a reliability of $+.98$, or nearly perfection. Computation shows, however, that even here there remains the equivalent of about half of one of the irrelevant factors of Table 42. As a matter of fact, theoretical absolute perfection in the reliability coefficient is never reached by any finite number of repetitions.

HOW TRUE RELATIONS MAY BE DEDUCED FROM FALLIBLE DATA

It is often of interest to know how well a fallible measure would correlate with a true measure of the function. It may be shown mathematically that this correlation is simply the square root of the reliability coefficient. Thus,

in the measure represented in Table 42, where the reliability coefficient is .67, a single repetition of the measure would correlate with a true measure to the extent of $\sqrt{.67}$, or .818. Exactly what this means in terms of concrete determining factors is shown in Table 44. Here the situation is exactly as shown in Table 42 except that variable (1) is represented merely by the true determiners, whereas the fallible measure (2) retains the seven irrelevant factors. As the above computations show, the determiners yield the same correlation (+.818) as the formula.

It should be noted that the various formulæ and tabular representations employed above assume that all repetitions of a measure, such as a test, are equally accurate. This is probably never strictly true. Occasionally it may be rather far from the truth. Successive repetitions of a test under conditions ordinarily considered as constant do of course tend to be of approximately equal value. However, of the first two repetitions of a test, the second is probably more likely to be the more accurate measure.

It is important to observe that successive repetitions of a test are not only not of equal accuracy, but that the function being tested may actually change its nature as the result of the testing. An excellent illustration of this is provided by an investigation by Ruch. About 50 subjects in the upper school grades were given the Binet-Simon Test. In addition, the subjects were given a card-sorting test which was repeated 50 times. The first five repetitions of the card-sorting test were correlated separately with the Binet. The remaining trials were first averaged by fives and then correlated. The obtained coefficients are shown in Table 45. An examination of this table shows a rather uniform progress from the fairly large coefficient of +.43 downward to a small but tolerably consistent negative value. This would seem to indicate that the early trials at card sorting

TABLE 44

SHOWING THE EFFECTS OF NUMEROUS SMALL IRRELEVANT FACTORS IN A FALLIBLE MEASURE ON ITS CORRELATION
WITH THE TRUE MEASURE (See Table 42)

VARIABLE	DETERMINERS					IRRELEVANT FACTORS													
	I	II	III	IV	V	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
True measure (1) .		2	1		3														
Fallible measure (2) .		2	1		3		1		1		1				1		1		1

$$\frac{4 + 1 + 9}{\sqrt{(4 + 1 + 9)(4 + 1 + 9 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1)}} = \frac{14}{\sqrt{294}} = +.818$$

TABLE 45

SHOWING THE CORRELATION OF SUCCESSIVE REPETITIONS OF THE CARD SORTING WITH BINET SCORE (Adapted from Ruch)

NO. OF REPETITIONS OF CARD-SORTING TEST	CORRELATION COEFFICIENT
1	+ .43
2	+ .40
3	+ .19
4	+ .22
5	+ .21
6-10	+ .06
11-15	+ .02
16-20	+ .04
21-25	+ .05
26-30	+ .01
31-35	- .02
36-40	- .08
41-45	- .04
46-50	- .06

depended in part upon some of the same determiners as the Binet, but that with practice these determiners were no longer involved in the activity.

From the foregoing it is evident that all correlations among aptitudes and tests are affected by attenuation. No correlation coefficient as obtained by direct computation ever exactly represents the relation existing between the true determiners of the variables in question, but is always distorted downward more or less by the ever present irrelevant factors. It is possible, however, to correct the obtained or "raw" coefficients in such a way as to make them approximate rather closely the true relations. Suppose that we have a fallible measure of an aptitude and a fallible measure of a test as represented in Table 46. The reliability coefficient of each variable may be computed readily by the usual method (see page 231), assuming that a repetition of either

TABLE 46

SHOWING HYPOTHETICAL DETERMINER RELATIONS BETWEEN A FALLIBLE APTITUDE CRITERION AND A FALLIBLE TEST

VARIABLES	TRUE DETERMINERS					IRRELEVANT FACTORS													
	I	II	III	IV	V	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Fallible aptitude criterion (0) .	4	1	3		5	2		1		2									
Fallible test (1) .		2	1		3		1		1		1		1		1		1		1

Reliability coefficient $r_{00} = +.85$ Observed $r_{01} = +.563$
 " " $r_{11} = +.67$ " " " " $r_{01} = +.747$

would give the same true determiner weights as shown, and a similar set of attenuating factors, but the latter completely uncorrelated with the first set. Such a situation for variable (1) has already been given in Table 42. By this method it may be shown that the reliability coefficient of (0) is $+.85$ and of (1) is $+.67$. By ordinary computation, as the variables stand, fallible test (1) correlates with fallible aptitude criterion (0) to the extent of $+.563$. A fourth computation based solely on the true determiners of the two variables (i.e., disregarding the irrelevant factors) shows that the "true" correlation of the variables is really $.747$. The source of these four correlations should be noted, especially that of the last two, as it reveals the significance of the technique of correcting "observed" correlation coefficients for attenuation.

It is obvious, of course, that the "true" r cannot be found in practical correlation work by direct computation from the determiners. This is because, with real experimental data, we know practically nothing concerning the actual determiners. The other three coefficients shown can be found, however, by direct computation from the original scores, provided independent measures are available of each variable. Fortunately, there is a formula by which the fourth or "true" coefficient may be calculated if the others are known. It is:

$$\text{"True" } r_{01} = \frac{\text{Observed } r_{01}}{\sqrt{r_{00}}\sqrt{r_{11}}} \quad . \quad . \quad . \quad (4)$$

Substituting the above values in the formula, we have:

$$\begin{aligned} \text{"True" } r_{01} &= \frac{.563}{\sqrt{.85}\sqrt{.667}} \\ &= \frac{.563}{.753} \\ &= +.747 \end{aligned}$$

Thus the theoretical "true" correlation coefficient, the found correlation corrected for attenuation, may readily be obtained.

In finding the "true" correlation under practical experimental conditions the procedure would be slightly different from that outlined above, which is the one usually presented in works on statistics. It will be recalled that two independent measures of each variable were required. This was to make the reliability coefficients possible. But only *one* of each of these measures was used in calculating the observed r (.563). In practice, correlations should always be computed on the basis of the most accurate measures available. *Statistical manipulation can never take the place of carefully obtained experimental data.* This means that the observed r should be computed on the basis of the *two combined measures* of each variable instead of only one. The reliability coefficients to be used with this r in the formula just considered would also be different. In this case the reliability coefficient would be the correlation of the two combined measures with two other combined measures of equal value, instead of a single measure with a single measure, as above. The reliability coefficient for the combined measures may be calculated from that secured from the single measures. The formula used is $\frac{2r}{1+r}$ which we have already had occasion to notice (page 236). The formula for the "true" r in this case then becomes:

$$\frac{r_{(0+0)(1+1)}}{\sqrt{\frac{2r_{00}}{1+r_{00}}} \sqrt{\frac{2r_{11}}{1+r_{11}}}} \dots \dots \dots (5)$$

THE TRUE VALIDITY OF A TEST BATTERY

The above formulæ which provide a complete correction for attenuation are of value chiefly in theoretical investigations. In such situations correlation coefficients are of almost no value unless fully corrected in this manner. In practical aptitude work, however, we have a somewhat different situation. A test battery may correlate with the criterion of an aptitude to a certain known extent. This coefficient is often taken as an indication of the "validity" of a test. Attenuation may, however, be assumed with certainty to have taken place both on the side of the criterion and of the tests. However, there is no practical value in learning how well the tests of a battery *would* correlate with the true aptitude *if* the tests were more reliable than in fact they are. But the situation on the part of the criterion is quite different. The tests are actually predicting the true criterion better than they predict the available fallible measure of it. But since we are interested in the thing itself rather than in any fallible measure of it, we desire to know how well our test scores correlate with the true aptitude. It thus comes about that in our desire to know how effective or "valid" our tests are we need to correct our coefficients on the side of the criterion alone. To do this there should be two independent measures of the criterion. These should be combined to form the criterion score against which the tests are correlated. Let the two criterion measures be 0 and 0 and the *predicted* criterion score be 1. Then the formula for the correlation between the prediction and the true criterion becomes

$$\frac{r_{(0+0)1}}{\sqrt{\frac{2 r_{00}}{1 + r_{00}}}} \dots \dots \dots (6)$$

Suppose, for example, a battery of aptitude tests correlates $+.55$ with a criterion score composed of two independent measures which correlate with each other to the extent of $.50$. Substituting these values in the above formula :

$$\frac{.55}{\sqrt{\frac{2 \times .50}{1 + .50}}} = \frac{.55}{\sqrt{.667}} = +.672$$

We find that the reliability of the combined criterion is $.667$ and that the battery correlates with the true criterion to the extent of $.672$. Thus $.67$ is found to be the true coefficient of validity instead of $.55$. It is important to note that in terms of actual forecasting efficiency (page 273) this represents a rise from 16.5 per cent to 25.8 per cent, or a gain of over 56 per cent in actual validity. The moral of this is that the validity of tests is always superior to that indicated by the raw correlation with the criterion measure, and often very much superior.

DETERMINERS OF DIFFERENT SIGN, LATENT AND NEGATIVE CORRELATION

We have seen how errors of measurement reduce correlation coefficients below the values which would have been obtained from wholly accurate measurements of the variables in question. These irrelevant factors act on a kind of dilution principle. Just as in the dilution of solutions some of the original solution always remains, so in correlation, no matter how much error of measurement takes place, there will usually remain some trace of correlation.

From the point of view of the theory of aptitude determiners it is of interest to observe that a situation is conceivable in which there may be a very high degree of commonality of determiners between two aptitudes, and no

TABLE 47

SHOWING PRESENCE OF STRONG LATENT CORRELATION, YET NO OBSERVABLE
CORRELATION WHATEVER

VARIABLE	DETERMINERS				
	I	II	III	IV	V
1	1		4		2
2	2		1		- 3

irrelevant or attenuating factors whatever, but in which there may be no correlation whatever between the aptitudes. The obscuring of the correlation in this case is due to one or more determiners in one variable having a weight of different sign from the corresponding determiners in the other variable. Such a situation is shown in Table 47. The application of the formula given on page 231 shows that these two variables would not correlate at all :

$$r_{12} = \frac{2 + 4 - 6}{\sqrt{1 + 16 + 4\sqrt{4 + 1 + 9}}} = .00$$

The apparent lack of correlation in the case under consideration is due to the negative weight of determiner No. V in variable 2. If this were positively weighted (or the corresponding weight of variable 1 were negatively weighted), the two variables would correlate $+.70$, which shows something of the actual closeness of relation between the two.

In case the importance of the determiners which differ in their signs is greater than the importance of the determiners of like sign, then there will result a negative correlation. Such a situation is shown in Table 48. Substituting in the usual formula, we have :

$$\frac{-2 + 4 - 9}{\sqrt{1 + 16 + 9\sqrt{4 + 1 + 9}}} = \frac{-7}{19.08} = -.367$$

TABLE 48

SHOWING A COMBINATION OF DETERMINERS WHICH WOULD PRODUCE A
NEGATIVE CORRELATION COEFFICIENT

VARIABLE	DETERMINERS				
	I	II	III	IV	V
1	- 1		4		3
2	2		1		- 3

The exact status of genuine negative correlation among human activities is somewhat uncertain. By "genuine negative correlation" is here meant, not those in which a negative sign is the result of peculiarities of scoring, but rather negative correlation which would result if both variables were scored in such a way that a large score would be a "good" score. When human activities are scored in this way, the great majority of those so far investigated are positive. It is rare indeed to find genuine negative correlations of any considerable size. In relatively simple motor activities, however, a considerable number of small negative correlations are found (see Table 35). No doubt many of these may be due to chance, the true functions being correlated practically zero. So many of such r 's are found, however, that it is difficult to believe that all are due to chance. The probability seems to be that most determiners are positive, but that a small number may be negative. The matter needs further investigation.

In seeking to gain a concrete notion of what a positive and negative weighting might be, we may cite the following purely hypothetical example. It has been suggested that "speed of decision" might be of positive value to a field officer in time of war, where prompt action is likely to be decisive; but that the same trait might be very undesir-

able and so have a negative value in scientific work, in which time is not so important a factor but in which precision, thoroughness, and maturity of judgment are most needful. Unfortunately there is no particular reason to believe that "speed of decision" is such a stable aggregate of determiners as to constitute a faculty such as would be required by the above example.

SPURIOUS CORRELATION

So far we have considered various determining mechanisms which tend to distort correlation coefficients downward. It is important to observe that distortions are by no means always downward. Extraneous factors may sometimes produce spurious correlation, exaggerating a correlation already existent or even causing one to appear where there would otherwise have been none at all. As an example representing a rather common type of spurious correlation, suppose the mental age were correlated with the size of the great toe in a group of children ranging in age from five to fifteen years. A rather large correlation would doubtless result. This would be produced by the irrelevant factor of age even though the behavior measured by the test were quite unrelated to the size of the foot. We may readily represent the situation as shown in Table 49. An inspection of this table shows that instead of the irrelevant factors being always different in each variable as in attenuation, they tend in dilation or spurious correlation to be the same in each variable.

Computing the various correlations from the determiners by means of the formula (page 231), we find them as follows:

$$r_{12} = +.4841$$

$$r_{13} = +.8728$$

$$r_{23} = +.5546$$

TABLE 49

ILLUSTRATING THE MECHANISM OF SPURIOUS CORRELATION. MENTAL AGE AND SIZE OF THE GREAT TOE ARE ASSUMED TO BE UNCORRELATED EXCEPT THROUGH THE INFLUENCE OF THE IRRELEVANT FACTOR OF AGE

VARIABLES	TRUE DETERMINERS				IRRELEVANT FACTORS		
	I	II	III	IV	I	II	III
Mental age (1)		2	1		4		
Size of great toe (2)				3	2		
Chronological age . . . (3)					3		

The spurious correlation is r_{12} , which amounts to $+.4841$. Yet an inspection of Table 49 shows that, except for the irrelevant factor of chronological age, these two variables are quite uncorrelated.

PARTIAL CORRELATION ILLUSTRATED

It sometimes happens that in cases such as that just considered it is possible to correct the obtained correlation coefficient for the distortion, provided the disturbing agent or factor is known and can be measured. The method of correction recalls that already employed in the correction for attenuation. In this case, however, it is called *partial correlation*. The formula is:

$$r_{12.3} = \frac{r_{12} - r_{13} \times r_{23}}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}} \quad . \quad . \quad . \quad (7)$$

where the only expression not familiar to the reader is that of $r_{12.3}$. This is read "correlation of 1 with 2 when variable 3 is kept constant." In the case of spurious correlation just considered it would mean the correlation of mental age with the size of the great toe, chronological age being constant. In other words, it gives the correlation that would be found between the mental age and the size of the great

toe if all the subjects were of exactly the same age. That the formula accomplishes this may readily be verified by substituting in the formula the various relevant r 's.

$$r_{12.3} = \frac{.4841 - .8728 \times .5546}{\sqrt{1 - .8728^2} \sqrt{1 - .5546^2}} = \frac{.4841 - .4841}{.4936 \times .8319} = .00$$

Thus $r_{12.3}$ is shown by the partial-correlation formula to be zero, exactly as represented in Table 49.

The ease with which we were able to penetrate the complex of factors involved in a correlational situation and remove the influence of a spurious factor must not mislead the reader into expecting too much from partial correlation. It might be supposed, for example, that the partial correlation technique would be able to reveal the details of latent correlation, as in Table 48, where none is shown by ordinary methods. It might even be thought an instrument for the isolation of the ultimate determiners of aptitudes. As a matter of fact, a number of investigations have been published which seem to have had substantially this point of view. It is not difficult to show, however, that any such hopes are without substantial foundation.

LIMITATIONS OF PARTIAL CORRELATION IN DETERMINER ANALYSIS

The fundamental fallacy ordinarily underlying the illegitimate use of partial correlation as a means of detecting the presence of hidden determining factors is the tacit assumption that, when one activity is held quantitatively constant by means of the formula, all the factors determining that activity are thereby held constant. This is in reality by no means the case except where, as in the example of the chronological age factor assumed above (Table 49), the variable held constant is strictly unitary. But where the variable which is held quantitatively constant is itself de-

pendent upon a number of independent determiners, as is usually the case in aptitude work, the partial-correlation formula holds constant only the joint *total* of the several potencies of the determiners. This means that the *separate* determiners are not held constant at all. On the contrary they are perfectly free to vary, often within a very considerable range. As a consequence of this, *results usually obtained by partial correlation have no significance whatever as revealing hidden determining factors.*

A simple example illustrating the possibilities and limitations of partial correlation as a means of isolating hidden determining factors in complex situations is given in Table 50. The various correlation coefficients are given

TABLE 50

SHOWING A HYPOTHETICAL SET OF DETERMINERS TO BE USED TO ILLUSTRATE THE MECHANISM OF PARTIAL CORRELATION

VARIABLES	DETERMINERS		
	I	II	III
1	1		
2		1	
3	1	1	

$$r_{12} = .00$$

$$r_{13} = +.7071$$

$$r_{23} = +.7071$$

beneath the table, computed directly from the determiners. If, now, we apply the partial-correlation formula to determine the correlation of variable 1 with variable 3 for constant values of variable 2 ($r_{13.2}$), the result is $+1.00$. This, of course, agrees with the naïve view. By holding 2 constant the formula has thereby eliminated the influence of determiner No. II which also constitutes that part of 3

which is not found in 1. What is left of 3 (determiner No. I) is identical with variable 1, and so variables 1 and 3 now correlate perfectly. The possibility of the above analysis was due to the fact that variable (2), the activity held constant, was a simple unitary variable.

But suppose now we apply the formula to determine the correlation of 1 with 2 for constant values of 3, which is a *complex variable* depending upon two independent determiners. The result is very different. Substituting appropriately in the partial-correlation formula, we have :

$$\begin{aligned} r_{12.3} &= \frac{.00 - .7071 \times .7071}{\sqrt{1 - .7071^2} \sqrt{1 - .7071^2}} \\ &= \frac{-.5}{\sqrt{.5} \sqrt{.5}} \\ &= -1.00 \end{aligned}$$

According to the naïve view the resulting value of -1.00 ought to represent a total negative correlation that somehow had been latent between 1 and 2 but now has been revealed by the partial-correlation technique. Moreover, this implies the existence of a negative determiner. The fact is, of course, that no suggestion of a negative determiner is to be found anywhere in Table 50, and that 1 and 2 have nothing at all in common. In this case, then, partial correlation clearly fails as an instrument of analysis for the isolation of causal determiners of behavior.

But what, then, is the significance of the partial-correlation value of -1.00 ? The answer to this question is revealing as to what partial correlation is and does. In the first place it must be recalled that partial correlation, by definition, gives the correlation which would be found between two variables which happen to be associated with constant values of a third variable. In the case under consideration we might wish to know how values of vari-

ables (1) and (2) would correlate with each other, assuming that they were also associated with a constant value of variable (3), such as 7. An inspection of Table 50 shows that variable (3) will always be the simple sum of the values of variable (1) and variable (2). This being the case, if variable (1) is 1, then variable (2) must be 6; and if variable (1) is 2, then variable (2) must be 5; and so on. In short, the larger one is, the smaller must be the other. This naturally produces a perfect negative correlation, as may be seen in columns (1) and (2) of Table 51, and exactly as indicated by the partial correlation formula.

TABLE 51

SHOWING THE RELATION OF THE VARIOUS VALUES OF VARIABLES (1) AND (2) OF TABLE 50 WHICH CORRESPOND TO A CONSTANT VALUE OF VARIABLE (3)

VARIABLE (1)	VARIABLE (2)	VARIABLE (3)
1	6	7
2	5	7
3	4	7
4	3	7
5	2	7
6	1	7

It is evident from the foregoing that as an instrument for the analysis of the factors producing highly composite results, partial correlation must be used with extreme caution. Apparently where the variable held constant is itself a composite, the results secured by partial correlation are essentially meaningless, so far as the ultimate determiners are concerned.

CHAPTER EIGHT

THE COMPOSITION AND YIELD OF TEST BATTERIES

IN the preparation of a series of tests from which to make up an aptitude battery, usually a considerably larger number of tests are tried out and correlated against the aptitude criterion than will finally be used. This is because rarely or never are all the tests thus tried out found to be suitable for use. Numerous practical considerations bearing on the choice of tests for the final test battery are to be given in a later chapter (pages 302 ff.). For the present we shall confine ourselves to the consideration of certain important principles bearing on composition and yield.

TWO GUIDING PRINCIPLES IN THE FINAL COMPOSITION OF TEST BATTERIES

In choosing the tests to make up a battery, two prime considerations must be observed:

(I) The tests should each correlate as highly with the aptitude criterion as possible.

(II) They should correlate as *low* with each other as possible.¹

Principle I is so obvious as hardly to need emphasis. Indeed, in the early days of mental testing it was customary to choose tests almost solely on this basis and to disregard the correlations among the tests themselves, thus ignoring principle II. In some cases a principle almost exactly opposite that of II appears to have been followed, the workers actually choosing tests which correlated as *highly* with each other as possible. Fortunately for the progress of aptitude testing,

¹ These principles hold for most practical purposes if the criterion and all tests are scored in such a way that a large score is a "good" score. For a detailed discussion involving the above principles but where both positive and negative correlation coefficients are involved, see pages 450 ff.

both principles I and II are now very generally observed by trained psychologists.

The folly of choosing for a battery two tests which correlate highly with each other is evident upon reflection. It is obvious that in so far as tests correlate with each other they are testing the same functions. When this correlation approaches the maximum, the two tests become identical, in which case to give both tests would be equivalent to giving the same test twice. If we assume that the test measured its function with reasonable accuracy the first time, it would be a simple waste of labor to measure it again by giving a second test almost identical in nature. The same principle holds, though to a lesser degree, as the correlation between tests decreases toward zero.

The operation of the two principles may become somewhat clearer from the consideration of a particular testing situation. We shall take the one presented in Table 52. Com-

TABLE 52

HYPOTHETICAL TESTING SITUATION TO ILLUSTRATE THE PRINCIPLES OF CHOOSING TESTS FOR A BATTERY

ACTIVITIES	DETERMINERS												
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
Aptitude (0)	2	4		2	3		1			1		3	
Test (1)		3	5	2							1		
Test (2)		3	4	1					2				
Test (3)					3	4	1				2		

putation from the determiners there represented shows the various correlation coefficients to be as follows :

$$r_{01} = +.40$$

$$r_{02} = +.39 \quad r_{12} = +.91$$

$$r_{03} = +.28 \quad r_{13} = +.06 \quad r_{23} = .00$$

According to principle I, test 1 is the best, since it correlates with the criterion to the extent of .40; test 2 is nearly as good, and test 3 is of least value. Clearly, if only a single test were to be chosen that test would be No. 1.

If two tests were to be chosen and combined into a battery, principle I considered by itself would indicate tests 1 and 2. But this would conflict with principle II because these tests have the very high intercorrelation of .91. As a matter of fact, appropriate computation shows that these two tests combined would correlate with the criterion to the extent of .404, which is only .004 better than test 1 alone. The combination of test 2 with 1 would evidently involve a mere waste of effort. The combination of tests 1 and 3, however, is evidently much more promising, since the two correlate only +.06. Computation in this case shows that the two combined would correlate with the criterion to the extent of +.475. This represents a decided advance, even though r_{03} is distinctly lower than r_{02} .

As a third possibility may be considered the combination of tests 2 and 3. This is quite in harmony with principle II, since r_{23} is .00, but in conflict with principle I, since it leaves out test 1, which shows the highest correlation with the criterion. Appropriate computation shows that the yield from this combination of tests would be +.481, or actually a trifle higher than the combination of 1 and 3. This is particularly striking, as it shows the theoretical possibility that the best single test (if standing alone) would not necessarily find a place on a battery. In practice, however, the rejection of the strongest single test would rarely take place. Ordinarily the strongest test will be taken as a nucleus or core around which the lesser tests are arranged in the most effective manner that the circumstances permit.

THE INCREASE IN YIELD CONTRIBUTED BY SUCCESSIVE TESTS
ADDED TO APTITUDE BATTERIES

We have seen how the two major principles of the choice of test batteries operate when applied on a small scale. We must now consider a number of interesting and important results which become apparent when these principles are considered systematically and on a larger scale. In order to simplify the problem we shall assume throughout the analysis (1) that every test correlates with the aptitude criterion exactly the same as every other test, and (2) that all the tests correlate with each other to exactly the same extent. It will accordingly be possible to let r' represent alike all correlations between tests and criterion and r'' represent alike all correlations between the tests themselves. Then if $r' = +.40$ and $r'' = +.20$, the various aptitude and test correlations of such a battery would be as follows:

$$\begin{array}{llllll} r_{01} = .40 & & & & & \\ r_{02} = .40 & r_{12} = .20 & & & & \\ r_{03} = .40 & r_{13} = .20 & r_{23} = .20 & & & \\ r_{04} = .40 & r_{14} = .20 & r_{24} = .20 & r_{34} = .20 & & \\ r_{05} = .40 & r_{15} = .20 & r_{25} = .20 & r_{35} = .20 & r_{45} = .20 & \end{array}$$

Subscript 0 as usual represents the criterion, and the numbers from 1 on represent the various tests.

With the testing situation thus simplified, analysis becomes relatively easy. It becomes possible, for example, to determine in a few moments the correlation with the criterion which any given number of such tests would yield. The formula by which this is accomplished is:¹

$$R = r' \sqrt{\frac{N}{1 + r''(N - 1)}} \cdot \cdot \cdot \cdot \quad (8)$$

¹ See 35 a.

where r' and r'' have the meanings given above, N is the number of such tests employed, and R is the correlation with the criterion of the combined battery when weighted in the best possible way by means of the multiple-regression equation.

Numerous questions concerning the yield of test batteries at once present themselves: How much does the addition of a second test to the first of a battery increase the yield over that given by the first alone? Do two tests, for example, yield twice as much as one? How much does a third test add to the yield of the first two? How much does the fifth or the seventh test add? How much greater is the yield from fifty or one hundred tests over that for five or six? By substituting appropriately in the above formula (No. 8) and assuming the fairly typical values $r' = +.40$ and $r'' = +.20$, we secure the series of yields shown in Table 53. The answers to our questions may now be obtained from the successive entries in this table by means of subtraction. In this way it is found that the second test of the battery adds .116 to the original correlation of .40. It will be noted that this, while considerable, is far from doubling the yield. The third test adds 6.8 points to the yield given by the first two, and so on. The amounts contributed by various successive additions to the battery are shown systematically in

TABLE 53

SHOWING THE YIELDS OBTAINABLE FROM VARYING NUMBERS OF TESTS
WHERE $r' = +.40$ AND $r'' = +.20$

2 tests, $R = .516$	50 tests, $R = .8596$
3 tests, $R = .584$	51 tests, $R = .8616$
4 tests, $R = .632$	100 tests, $R = .8768$
5 tests, $R = .667$	101 tests, $R = .8772$
6 tests, $R = .693$	1000 tests, $R = .8926$

TABLE 54

SHOWING THE AMOUNT ADDED TO THE YIELD (R) BY EACH SUCCESSIVE TEST ADDED TO THE BATTERY, AS DERIVED BY SUBTRACTION FROM TABLE 53

No. 1, .40	No. 5, .035
No. 2, .116	No. 6, .026
No. 3, .068	No. 51, .002
No. 4, .048	No. 101, .0004

Table 54. The facts of Tables 53 and 54 (up to where $N = 20$) are shown graphically by the curve marked “.40, .20” in Figure 34 on the next page.

For purposes of comparison there is plotted in Figure 34 a companion curve based on the relation “.40, .05” and so marked. It will be observed that the latter, while originating at the same point, rises very much more steeply, indicating a markedly higher degree of combining efficiency. This is due to the fact that in this curve the tests correlate among themselves .05, whereas in the former curve, representing a less efficient combination, the tests correlate among themselves +.20. Another set of comparisons is presented in the series of three curves originating at .30. Of these “.30, .05” rises to perfection at about test 22, “.30, .20” attains a yield of .61 after some twenty tests, and “.30, .50” reaches only .41 with the same number of tests. Once more these curves illustrate in a most striking manner the desirability of having the tests of a battery correlate as low with each other as possible.

The crossing of curves “.30, .05” and “.40, .20” at about test 10 illustrates the interesting fact that with less than ten tests in a battery the combination of “.40, .20” is the more effective, whereas in a battery of more than ten tests the combination “.30, .05” is decidedly to be preferred.

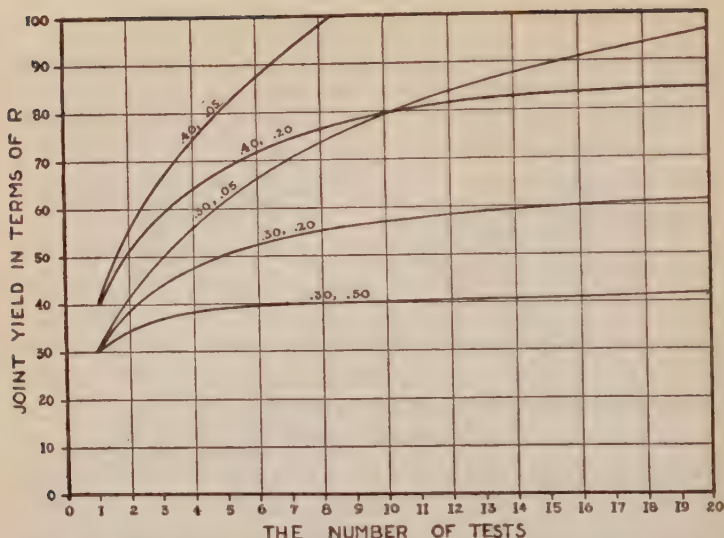


FIG. 34. Showing the joint correlation yield (R) of varying numbers of tests. All tests contributing to the yield represented by any given curve are assumed to be statistically alike; i.e., all have the same correlation with the criterion and the same correlation with each other. The correlations in question are shown above each curve. Thus, .30, .20 means that all tests making up that particular curve correlate $+.30$ with the criterion and $+.20$ with each other. (See Figure 35.)

This situation is illustrative of a rather general principle that for a small number of tests a high r' accompanied by a high r'' is to be preferred to a low r' accompanied by a low r'' . But with a large number of tests the situation is likely to be reversed.

PRINCIPLE OF DIMINISHING RETURNS IN THE SIZE OF TEST BATTERIES

Perhaps the most striking fact revealed by the curves of Figure 34 is the radical tendency to diminishing returns as successive tests are added to the battery. This is shown

in the tendency of the curves to flatten out with the continuous addition of tests. This tendency is probably found in every curve representing actual aptitude situations. The curves show, moreover, that the tendency to diminishing returns is greatly increased with the size of r'' , which is in harmony with principle II (page 254) governing the final choice of tests for a battery.

The facts concerning the yield of test batteries considered above have far-reaching implications as to the practical possibilities in the way of aptitude testing. It is a fact well known to aptitude psychologists that tests for genuine vocations rarely yield r' values much above .40. Usually they are lower. On the other hand, the r'' values are likely to range considerably above +.20. If the tests are chosen wisely, however, it is probable that in most cases the r'' values can be kept down to near this figure. Unfortunately, so little research has been devoted specifically to this subject that little is known about what would be accomplished by a vigorous effort in the way of finding low r'' values. In any case, ".40, .20" is a very favorable combination as correlations run in present practice. In addition, we have assumed the possibility of finding a large number of different tests of the same high degree of excellence, which is itself optimistic to the point of extravagance. With all these favorable conditions it is somewhat startling to observe what a small amount of increase in correlation yield is produced by the addition of another test beyond six or seven. The tenth test of a ".40, .20" battery adds only 1.2 points to the correlation yield. The question inevitably arises whether an increase of a single point or so in the correlation yield is worth the extra time and labor involved in giving and scoring an entire additional test unit. In any case it is perfectly obvious that because of this law of diminishing returns a place must be reached sooner or later where the addition of a new test

will not contribute enough to the prognostic value of the battery to justify the incidental expense involved. This fact alone must inevitably limit the size of test batteries. It is a matter of considerable importance to aptitude psychology whether this critical point is reached before or after a really satisfactory *R* is secured.

It is evident that the number of tests which may be included before the point is reached where the expense exceeds the return is dependent upon a number of factors. One of these which has received very little theoretical consideration is the net cost of purchase and administration of the test units. It is perfectly clear, for example, that when batteries made up of tests of the ".40, .20" variety increase in number of units much beyond ten, the costs associated with the addition of each successive test would have to be very small indeed to be justified. The ultimate fate of the aptitude-testing movement may thus depend quite as much upon the possibilities of reducing testing costs as upon the prognostic potencies of the test units devised.

TESTS WHICH CAN NEVER MAKE A PERFECT BATTERY REGARDLESS OF THE NUMBER EMPLOYED

It is often a source of surprise, even to persons with considerable experience with tests, to learn that it would ordinarily be impossible to attain a perfect prediction by means of a test battery, even though an indefinitely large number of tests of a given variety were available. This state of affairs is suggested in Figure 34 by the extreme tendency to diminishing returns in certain of the curves, such as ".30, .20" and ".30, .50." It is true that in certain theoretical situations the battery attains perfect prediction very promptly, but these conditions are probably never encountered in aptitude practice. Whether or not a particular combination of r' and r'' values such as assumed above will attain perfec-

tion may be determined rather readily. The rule is that for R to reach perfection r'' must not exceed the square of r' . Accordingly, ".40, .05" easily reaches perfection, as shown in Figure 34. But the combination ".40, .20" can never reach perfect prediction, no matter how many tests may be used, because .20 is larger than .40², which is only .16.

In cases where the maximum yield falls short of 1.00, its amount may be determined by the formula:

$$R_{N=\infty} = \sqrt{\frac{r'^2}{r''}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Applying this formula to the ".40, .20" combination, we have, for example,

$$R_{N=\infty} = \sqrt{\frac{.40^2}{.20}} = .8944$$

As might have been expected, this does not differ greatly from the yield when $n = 1000$, as shown in Table 53.

INDEPENDENT TESTS AS FRACTIONAL PARTS OF PERFECTLY PREDICTING BATTERY

A special case involving the yield of test batteries where $r'' = .00$ has some significance as to the nature and meaning of the correlation coefficient. If in Formula 8 (page 257), r'' be given the value of zero, we have:

$$r' = \frac{R}{\sqrt{N}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

By means of this we may readily determine the correlation which independent tests would need to have with the criterion for any given number of them to combine to produce a perfect prediction. A typical series of such values is given in Table 55. From the first entry it is evident that two entirely independent tests would each need to correlate .707 with the criterion in order that when combined they should corre-

TABLE 55

SHOWING THE CORRELATION VARYING NUMBERS OF STRICTLY INDEPENDENT TESTS MUST HAVE WITH THE CRITERION IN ORDER TO PRODUCE A PERFECT PREDICTION

$n = 2,$	$r' = .707$
$n = 3,$	$r' = .577$
$n = 4,$	$r' = .50$
$n = 5,$	$r' = .447$
$n = 10,$	$r' = .316$
$n = 100,$	$r' = .10$

late with it perfectly. Since, in a sense, each represents half of the correlation, one may say that the half of a perfect correlation is not .50, but .707. In a similar manner, a third of a perfect correlation is not $.33\frac{1}{3}$, but .577; a fifth is not .20, but .447; and so on.

Or we may reverse the statement and say that an independent test correlating .707 with a criterion will constitute 50 per cent of a perfect battery, one of .447 will constitute 20 per cent of a perfect battery, and so on. A systematic presentation of the testing situation from this point of view is shown in Table 56.

COEFFICIENT OF MULTIPLE CORRELATION NOT A PERFECT INDEX OF BATTERY EFFICIENCY

Up to this point in our discussion of the joint returns from tests combined into batteries, we have used the total correlation (R) as the index of yield. In this we have followed the general practice among psychologists. Unfortunately, the correlation coefficient is not very satisfactory as such an index. This is because its relation to the true yield is somewhat indirect and by no means simple. Mathematically the correlation coefficient is a trigonometric function. It is the natural tangent of an angle made by the intersection of certain lines drawn through a contingency table upon which the arrays of correlated values are distributed (105).

TABLE 56

SHOWING THE FRACTIONAL PART OF A PERFECT BATTERY MADE UP BY SINGLE INDEPENDENT TESTS CORRELATING VARIOUS AMOUNTS WITH THE CRITERION

CORRELATION WITH CRITERION	PART OF PERFECTLY PREDICTING BATTERY
.00	.0%
.05	.25%
.10	1%
.15	2%
.20	4%
.25	6%
.30	9%
.35	12%
.40	16%
.45	20%
.50	25%
.55	30%
.60	36%
.65	42%
.70	49%
.75	56%
.80	64%
.85	72%
.90	81%
.95	90%
1.00	100%

For the psychologist to whom mathematics is usually a means rather than an end, this trigonometric aspect of the test-aptitude relation is not very illuminating. As a matter of fact, there is abundant evidence that it has actually resulted in gross and widespread misconceptions as to the power and value of psychological tests.

Perhaps the chief reasons for this misunderstanding are that the correlation coefficient takes the form of a decimal and that its values range from .00 to 1.00. It thus has a striking but deceptive resemblance to the universally employed

percentage system of representing efficiencies. As a consequence, a test correlating .40 with an aptitude has too often been thought to be 40 per cent perfect in forecasting efficiency, and a battery correlating .70 with an aptitude as 70 per cent efficient. A more serious error could scarcely be made. Actually, a correlation of .40 corresponds to less than 9 per cent forecasting efficiency, and a correlation of .70 corresponds to less than 29 per cent efficiency (page 273).

“PER CENT OF SUBJECTS RIGHTLY PLACED” AN AMBIGUOUS
INDEX OF TEST-BATTERY EFFICIENCY

A second method of interpreting the correlation between a test battery and an aptitude is to state the per cent of subjects who can be rightly placed in the aptitude by means of the test scores. This, while possessing a certain amount of superficial plausibility, is in reality very little better than the one just mentioned. The main reason is that the per cent of subjects rightly placed (i.e., within half a point of where predicted) varies not only with the size of the R (which is to be expected) but also with the coarseness of the scale used to measure the criterion (which is ordinarily overlooked). The coarseness of the scale unit is obviously irrelevant to the potency of the test. As an example of the ambiguities to which this leads, an R of .60 will place 20 per cent within a half point of their true aptitude on a 10-point scale; 38 per cent on a 5-point scale of A, B, C, D, and E, such as is commonly used in academic marking; and 59 per cent on a 3-point scale such as *excellent*, *good*, and *poor* (Table 57). A second serious defect is that an R of zero does not represent a zero per cent of subjects rightly placed. Instead, it places on a 10-point scale, 16 per cent correctly; on a 5-point scale, 31 per cent; and on a 3-point scale, 50 per cent. As a striking example of the inadequacy of this method of representing forecasting efficiencies of a test, it may be noted

TABLE 57

SHOWING FOR VARIOUS DEGREES OF CORRELATION AND COARSENESS OF CRITERION SCALE THE NUMBER OF INDIVIDUALS ACTUALLY FALLING WITHIN $\frac{1}{2}$ POINT OF WHERE PREDICTED. THEORETICALLY PERFECT DISTRIBUTIONS ARE ASSUMED.

<i>R</i>	COARSENESS OF CRITERION SCALE		
	10 Points	5 [Points]	3 Points
.00	16%	31%	50%
.40	17%	34%	53%
.50	18%	35%	56%
.60	20%	38%	59%
.70	22%	42%	65%
.80	26%	50%	73%
.90	35%	64%	87%
.95	48%	80%	97%

that an *R* of zero will place more of the subjects rightly on a 3-point scale than an *R* of .95 will place rightly on a 10-point scale. The various numbers placed correctly on scales of the three degrees of coarseness are shown systematically in Table 57.

It may be added that when the attempt is made to state the forecasting efficiency of a test or battery in terms of the per cent of subjects correctly placed by *rank*, the situation is even worse, if possible. In such an attempt we encounter all the difficulties just enumerated in connection with the per cent rightly placed on an ordinary point scale, and in addition there result complications arising from the fact that ranks are not true units but represent extremely varying intervals on a true scale (Fig. 54, page 384). It thus comes about that for a single value of *R* the number correctly placed by rank will vary not only for every different size of group ranked, but also for the different parts of the distribution of any single ranked group.

THE INDEX OF FORECASTING EFFICIENCY (*E*)

The ultimate purpose of using aptitude tests is to estimate or forecast aptitudes from test scores. A test which does this with slight error is a good test. A test which does this with much error is a poor test. The real forecasting efficiency of a test or battery is therefore at bottom inversely proportional to the amount of error resulting when the test is used to forecast an aptitude criterion. The basic notion is accordingly very simple. If the error made by a battery correlating zero with a criterion (i.e., a battery of no efficiency whatever) were 16 points and the error made by a mediocre battery were 12 points, the mediocre battery would clearly reduce the forecasting error by 4 points out of a possible 16. The test is therefore, in a perfectly simple and direct manner, 25 per cent efficient. Similarly, if the forecasting error made by a battery in predicting the same criterion were 14.4 points, then this battery would serve to reduce the forecasting error by 1.6 points out of a possible 16. This latter battery would therefore be 10 per cent efficient.

Fortunately, the simple and natural percentage concept of test efficiency suggested above is closely related to the correlation coefficient with which we are so familiar. When the correlation is known, the forecasting efficiency of a battery may be calculated at once. The formula is :¹

$$E = 1 - \sqrt{1 - R^2} \quad . \quad . \quad . \quad . \quad (11)$$

In the formula, *E* represents the per cent of perfect forecasting efficiency of a test battery in predicting its aptitude criterion. For example, if we have a test battery correlating .7288 with a criterion, the true forecasting efficiency may be secured by the formula as follows :

¹ See 42, page 76; 35, page 596; 39, page 32.

$$\begin{aligned}
 E &= 1 - \sqrt{1 - .7288^2} \\
 &= 1 - .685 \\
 &= .315
 \end{aligned}$$

The battery in question therefore has a forecasting efficiency of 31.5 per cent.

That E really constitutes a true index of the forecasting power of a test battery may readily be verified by means of the miniature set of data shown in Table 58.

TABLE 58

A MINIATURE SET OF DATA SHOWING A SET OF CRITERION SCORES AND THREE SETS OF TEST SCORES

SUBJECT NO.	CRITERION (X_0)	TEST A (X_1)	TEST B (X_2)	TEST C (X_3)
1	13	9	0	14
2	2	8	2	8
3	4	5	3	5
4	6	7	4	2
5	10	11	6	6
Mean	7	8	3	7
σ	4	2	2	4

The multiple-regression equation by means of which this criterion may be most accurately estimated from the three sets of test scores is:

$$\bar{X}_0 = .5 X_1 + .75 X_2 + .75 X_3 - 4.5^1$$

Substituting the test scores of each subject in turn, we obtain the estimates or "predictions." In the case of subject No. 1 this becomes:

$$\begin{aligned}
 \bar{X}_0 &= .5 \times 9 + .75 \times 0 + .75 \times 14 - 4.5 \\
 &= 4.5 + 0 + 10.5 - 4.5 \\
 &= 10.5
 \end{aligned}$$

¹ The derivation of this equation is given in full detail on pages 457 ff.

TABLE 59

GIVING THE ERROR OF ESTIMATE OF TEST BATTERY SHOWN IN TABLE 58
WHICH CORRELATES .7288 WITH ITS CRITERION

For purposes of comparison there is given a parallel set of figures based on the errors of estimate of a hypothetical battery correlating zero with the criterion.

		BATTERY SHOWN IN TABLE 58 CORRELATING .7288 WITH CRITERION			HYPOTHETICAL BATTERY CORRELATING ZERO WITH CRITERION		
1	2	3	4	5	6	7	8
Subject No.	Criterion ($\bar{X}_0$)	Criterion "Forecast" ($\bar{X}_0$)	Error of "Forecast"	Error Squared	Criterion "Forecast" ($\bar{X}_0$)	Error of "Forecast"	Error Squared
1	13	10.5	2.5	6.25	7	6	36
2	2	7.0	5.0	25.0	7	5	25
3	4	4.0	.0	.0	7	3	9
4	6	3.5	2.5	6.25	7	1	1
5	10	10.0	.0	.0	7	3	9
Means	7	7.0	2.0	7.5	7	3.6	16
Square root of means					2.7386		
					4		

These estimates or "forecasts" for each subject are given in column 3 of Table 59. The corresponding estimates of the hypothetical battery correlating zero with the criterion are given in column 6. These latter values are all alike because multiple-regression equations with zero R 's always predict at the mean of the criterion, no matter what the test scores.

The error made in the prediction of subject No. 1 is 2.5 points, of subject No. 2 is 5 points, and so on. These errors of "forecast" (or *estimate*, as they are frequently called) are given in column 4 for the real battery and in column 7 for the hypothetical battery of zero R . The latter averages 3.6

points of error, whereas the former averages only 2.0 points. From these results an approximation to the forecasting efficiency might be computed directly. It is customary, however, for statisticians to use for this purpose the standard error of estimate rather than the mean values just mentioned. In the long run both methods give the same general relative results, but the standard error of estimate is regarded as more accurate. The errors in each column are therefore squared, the squares being recorded in the following column. These are added and averaged, and the square root of these averages is found. In the case of the real forecasts this standard error of forecast is 2.7386, whereas for the hypothetical battery of no value it is 4. Our battery has therefore reduced the standard error of estimate

$$4 - 2.7386$$

or 1.2614 points. Dividing this by 4, we have

$$1.2614 \div 4 = .315 \text{ or } 31.5 \text{ per cent}$$

Thus the percentage of forecasting efficiency derived from the formula for E is exactly verified from the errors themselves.

THE INCREASE IN FORECASTING EFFICIENCY CONTRIBUTED BY SUCCESSIVE TESTS

With the aid of E , the basic index of forecasting efficiency, we return to the consideration of the problem of diminishing returns in the yield of test batteries as the test units increase in number. This is best shown by means of a series of curves (Fig. 35). In order to facilitate comparisons, exactly the same r' and r'' combinations have been used as were employed in Figure 34. It is evident from Figure 35 that the phenomenon of diminishing returns, while still conspicuous with the correlation combinations where r'' is fairly large, has disap-

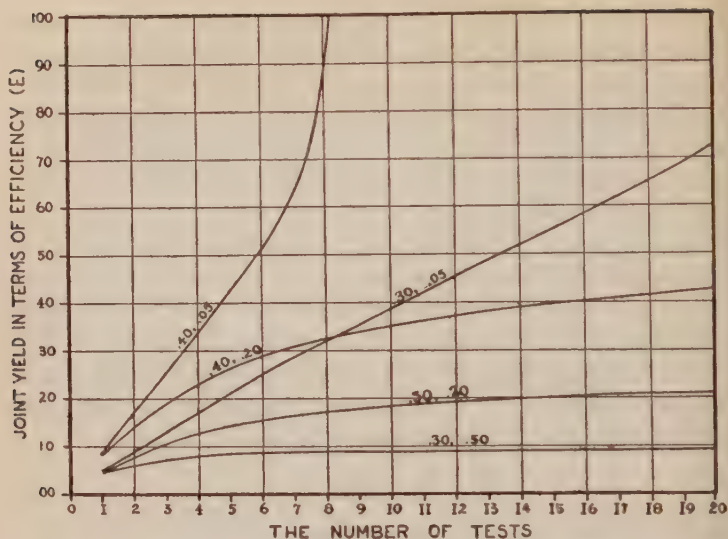


FIG. 35. Showing the joint efficiency yield (E) of varying numbers of tests. As in Figure 34, all tests contributing to the yield represented by any given curve are assumed to be statistically alike; i.e., all have the same correlation with the criterion and the same correlation with each other. These correlations are shown above each curve. They are exactly the same as those made the basis of the curves of Figure 34, so as to facilitate comparison between yields in terms of R and E .

peared in certain other cases. The combination “.30, .05” is striking in that the line is almost perfectly straight. The combination “.40, .05,” on the other hand, has actually reversed the form of its curve, now showing a condition of increasing returns.

In order still further to compare Figures 34 and 35, it will be well to trace out a typical correlation combination (.40, .20) as represented by the two methods. It will be observed that at the beginning the E -value is conspicuously lower, being only about a fifth the size of R , whereas at 20 tests it is fully half the size of R . Between test 8 and test 16

E increases a full third, whereas R increases less than a tenth. The implication of this and similar comparisons for testing practice is that the later tests of a battery contribute relatively much more to the yield, as compared to the early tests, than is indicated by R .

RELATION OF THE CORRELATION COEFFICIENT TO PER CENT
OF FORECASTING EFFICIENCY

The yield of aptitude batteries is at present almost universally reported in terms of the correlation coefficient. But they must be judged, apart from the cost of administration, upon the basis of their actual efficiency in forecasting. It therefore becomes a matter of considerable importance not only to have thorough understanding of the relation between the two, but also to be able readily to translate correlation values into equivalent forecasting-efficiency values. Such a series of equivalent values is given in Table 60.

Perhaps the most striking point about this table is the

TABLE 60

SHOWING THE RELATION OF THE CORRELATION COEFFICIENT (R) TO THE
PER CENT OF FORECASTING EFFICIENCY (E)

R	E
.10	$\frac{1}{2}\%$
.20	2%
.30	5%
.40	8%
.50	13%
.60	20%
.70	29%
.80	40%
.90	56%
.95	69%
.98	80%
1.00	100%

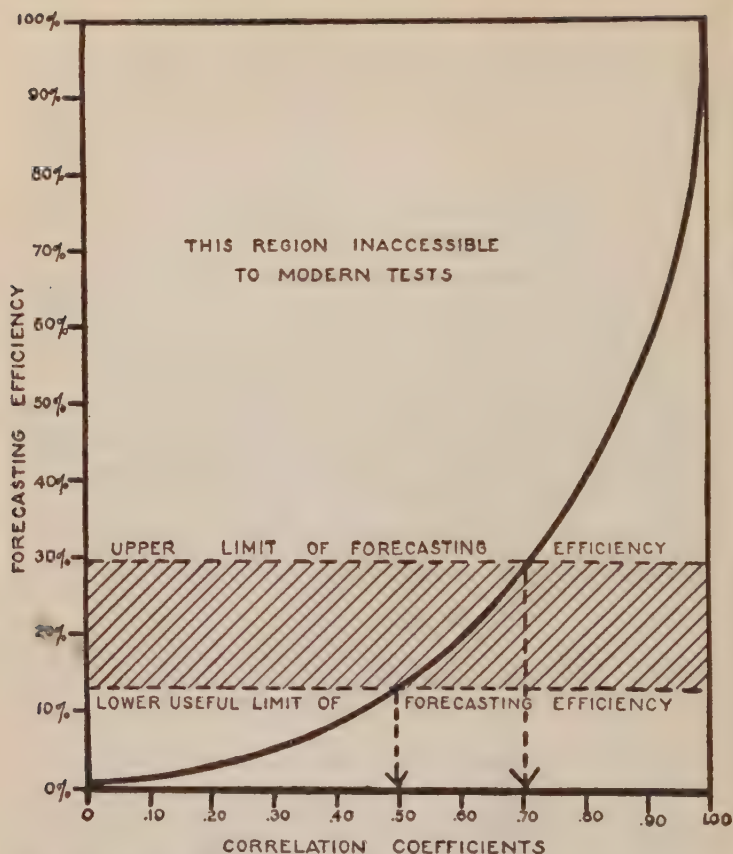


FIG. 36. Graphic representation of the relation between the correlation coefficient and forecasting efficiency. The range of useful forecasting efficiency of modern aptitude test batteries is indicated by the shaded area. It is assumed here that the correlation coefficients have been corrected for reductions in their size due to inaccurate measurement of the criterion. This is generally considerable.

remarkably small forecasting efficiencies corresponding to R values below .50. The same general tendency may be seen in the fact that the forecasting efficiency arises as much (20 points) between correlations .98 and 1.00 as between correlations .00 and .60! It is important to observe that *this zone of extremely low forecasting efficiency is exactly the zone where practically all modern aptitude correlations fall.* To the enthusiastic user of tests this may be somewhat disappointing, but the sooner these facts are fully realized the better for all.

The low forecasting efficiency of modern tests as shown by Table 60 inevitably raises the question as to their practical range of usefulness. Under ordinary conditions it seems likely that tests having a forecasting efficiency below about 10 per cent would not be worth using. This means that batteries correlating below .45 or .50 with a true criterion will hardly be useful unless they are extremely inexpensive. At the other extreme, correlations on genuine aptitudes running above .70 or .75 are so rare at present as to be practically non-existent. This means that the practical range of forecasting efficiency of modern psychological tests occupies the narrow zone roughly between 10 and 30 per cent. The situation is represented graphically in Figure 36, the curve showing the relation between R and E and the shaded portion the zone of practical forecasting efficiency of modern test batteries.

In connection with the practical yield from test batteries, the question is often asked, "What is a high and what is a low correlation?" Very different answers have been given at one time or another. In the light of the facts presented above we may make a number of fairly definite statements. These may be summarized as follows:

Below .45 or .50, practically useless for differential prognosis.¹

From .50 to .60, of some value

From .60 to .70, of considerable value

From .70 to .80, of decided value but rarely found

Above .80, not obtained by present methods

WHY THE YIELD FROM TEST BATTERIES IS ALWAYS LIMITED

In view of the low forecasting efficiency of aptitude tests, the question as to its cause naturally arises. The answer has considerable significance, not only for the present but for the future possibilities of aptitude testing as well.

If we could isolate and measure separately each determiner of an aptitude, it would be possible by means of suitable weighting secured by the multiple-regression equation to make predictions of 100 per cent efficiency. But we are not able to do this. Instead, *our tests are themselves aggregates of determiners which are already weighted by nature* very much like the aptitude itself. Thus in Table 61, Test 1 is based on determiners III, IV, V, and X, which are respectively

¹ This statement assumes that the correlation coefficients have been corrected for attenuation due to errors in measuring the criterion (pages 231 ff.). It is a well-known fact that the ordinary correlation coefficients obtained in practice are really lower than the actual correlations between the tests and the true aptitudes.

It should also be added that tests showing somewhat lower correlations than .50 may be useful in selecting individuals for a given purpose such as in employment testing or university entrance examinations, where considerable numbers of individuals falling below an arbitrary critical test score may be rejected. In general the higher the percentage of rejections that is permitted, the lower the useful correlation limit. Thus, in a private communication, Terman states that if the critical score of rejection on the Thorndike College Entrance Test (which ordinarily correlates about .50 with university marks) be placed at about 90, the number of scholastic failures would be reduced nearly to zero. This would involve, of course, the rejection of many fairly good students along with the very weak ones. But in employment work and in certain universities where large numbers of applicants must be turned away in any case, this works no particular hardship to the individual and yet results in a distinctly superior level of aptitude among those individuals finally selected.

TABLE 61

SHOWING A TYPICAL TEST-APTITUDE SITUATION IN TERMS OF DETERMINERS. EVERY DETERMINER OF THE APTITUDE IS TOUCHED BY ONE OR MORE OF THE TESTS, YET THE BATTERY YIELDS AN R OF ONLY .53.

	DETERMINERS												
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
Aptitude (0)	2		4		2	3		1			1		3
Test (1)			3	4	1					2			
Test (2)						3	4	1				2	
Test (3)	5	3	1				2						
Test (4)				3					4		2		1

weighted by the nature of the behavior involved: 3, 4, 1, and 2. No possible weighting of the obtained scores of this test by a psychologist can seriously disturb this natural weighting. The weighting secured by means of the multiple-regression equation is accordingly only a secondary affair, since it can apply only to *groups* of determiners as found in the tests. It thus comes about that even though one or another of our tests may touch every determiner or factor involved in an aptitude, the correlation between the aptitude and the test battery may be far from perfect and its forecasting efficiency may be actually so low as hardly to be of practical utility.

The reality of the conditions described above may be illustrated by the specific test-aptitude situation shown in Table 61. It will easily be observed by inspection that every single determiner found in the aptitude is found in the test battery. One determiner (No. III) is found in two different tests.

The various correlation coefficients as computed from the determiners by formula, page 231, are as follows:

$$r_{01} = + .385$$

$$r_{02} = + .275 \quad r_{12} = .00$$

$$r_{03} = + .338 \quad r_{13} = + .087 \quad r_{23} = + .233$$

$$r_{04} = + .137 \quad r_{14} = + .400 \quad r_{24} = .00 \quad r_{34} = .00$$

These correlation coefficients are fairly typical of those secured from ordinary testing operations. Appropriate computations reveal the fact that if the tests represented in Table 61 were weighted in the best possible manner by means of a multiple-regression equation,¹ the correlation yield (R) would be only .53. This corresponds to an actual efficiency of only 15 per cent. This is a matter of very considerable significance for the future of aptitude testing. It means that we may have involved in our test battery absolutely all of the determiners of the aptitude and yet the resulting battery may be almost without value.

¹ It is also a matter of considerable theoretical interest to observe that despite the fact that no negative determiner is to be found in Table 61, the regression equation contains a negative weight for test No. 4. Assuming all means and standard deviations to be unity, the equation is:

$$\bar{X}_0 = .171 + .376 X_1 + .215 X_2 + .256 X_3 - .009 X_4$$

The moral of this is, of course, that the regression equation, just as partial correlation (pages 250 ff.), may be an utterly misleading instrument for purposes of theoretical analysis of aptitude determiners. But if the multiple-regression equation is regarded as primarily a practical estimating or forecasting device, all of the mystery of this negative weight vanishes. It means that test No. 4 can contribute more to the yield of the battery as a whole by receiving a negative weight. An inspection of Table 61 suggests that this probably results from the fact that by weighting test No. 4 negatively there will result a tendency to neutralize in test No. 1 the influence of the irrelevant determiner IV, since this determiner appears prominently in both tests. Clearly test No. 1 would be much stronger with the influence of this irrelevant determiner eliminated.

PART TWO

METHODS

CHAPTER NINE

THE PSYCHOLOGICAL ANALYSIS OF OCCUPATIONAL BEHAVIOR

It has already been pointed out above (pages 257 ff.) that single tests are rarely or never adequate for detecting latent aptitudes. This is probably because aptitudes are ordinarily made up of a complex of abilities only a few of which are measured by any one test. As a result of this very general principle, the central problem of modern aptitude testing is one of devising test batteries. And a modern test battery is not merely a random aggregation of tests, the various scores of which are combined by simple addition or averaging. Such amateurish methods generally sacrifice by their clumsy technique the critical margin of efficiency which makes the difference between a useful battery and one which will not repay the expense of giving. While nearly universal in the early days of aptitude testing, these methods are now being replaced rapidly by more scientific procedures. Accordingly, we shall first sketch in broad outline the methods by which one goes about constructing a scientific battery of tests. There will then be taken up in succession for detailed consideration the various methods and special procedures which are necessary to carry through the different stages of the process, a chapter being devoted to each. It is intended that these chapters on method shall be sufficiently explicit and detailed to constitute a useful guide for those wishing to construct modern aptitude-test batteries.

GENERAL OUTLINE OF THE SIX STEPS OF TEST-BATTERY CONSTRUCTION

The first step in the construction of a battery of aptitude tests is to make a careful psychological analysis of the activity

or vocation in question. The purpose is to discover, in so far as this is possible, what traits or characteristics of human behavior lead to success or failure in this vocation. It is to be noted that by "psychological analysis" is not meant a mere armchair cogitation in vague general terms as to the traits which one might imagine as essential for success. What is here meant is a more or less protracted, objective, and systematic study of the behavior of individuals actually engaged in the particular activity. To this end there should be available for observation and study, while at work, a considerable number of people in various stages of training who show a wide range of natural aptitude in the task. The detail with which this psychological analysis of the aptitude should be carried out naturally will vary according to the amount of time available for the project as a whole.

The second step is the choice of a preliminary battery of tests which shall measure as well as possible the various pivotal traits emerging from the aptitude analysis as probably significant. To do this adequately there is required a wide knowledge of the complex psychology of the most varied types of tests and what they measure as shown by experimental trial in previous investigations. Here, again, mere armchair speculation as to whether a test will tap a significant part of a previously uninvestigated aptitude is highly unreliable. For this reason it is usually necessary, even with the best available information at hand, to choose a field of tests two or three times as large as that desired for the final battery, so that the tests representing incorrect guesses may be eliminated and still leave enough successful ones to make up a satisfactory battery.

The next thing necessary is to try out the preliminary field of tests to determine *objectively* which are to be preserved in the final battery and which are to be discarded. This part of the process may be called *testing the tests*. It bulks very large

in an aptitude-testing project, making up steps three, four, and five of the total process. Step three is the administration of the preliminary battery of tests to a large number of individuals who are about to start training in the aptitude under investigation, but who have not as yet had any actual experience in it.

The fourth step of the process presents by all odds the most serious difficulties and requires the greatest ingenuity and resourcefulness on the part of the psychologist. This is the securing of a quantitative determination of the final aptitudes or vocation proficiencies of the trial group of subjects just mentioned, after they have finished their training. When this quantitative determination is secured, it is called the *criterion score*. The difficulty in securing this criterion score lies in part in the fact that as a rule the individuals in the group are not working on exactly the same materials, machines, or tasks, so that the outputs of the various individuals are not comparable. A still more serious difficulty is that in a large number of vocations there is no objective unit for measuring output. A concrete example of this is found in the difficulty of determining the relative excellence of a series of free-hand drawings made by different persons, even though all the drawings may be based on the same object. Fortunately this problem of the criterion score is gradually yielding to the methods of modern experimental psychology.

The fifth step is to check the test scores of the trial subjects which were secured by step three, against their criterion scores obtained by step four. This is done by the computation of correlation coefficients. A careful examination of the coefficients thus obtained usually reveals a half or more of the tests as so feebly related to the criterion that they will contribute little or nothing to the joint yield of the battery. Also it frequently happens that two or more tests which correlate fairly well with the criterion will be found to sample

nearly the same aspect of the criterion. Since there is no value in testing the same function more than once, even though it may be an important one, all but the best of such duplicating tests will be discarded. The tests which remain after this ruthless try-out will make up the final test battery.

The sixth and final step in the process of developing an aptitude-test battery is the determination of the relative weights to be given the various surviving tests. It is obvious that the particular element or constituent part of an aptitude sampled by one test will rarely be exactly as important as that sampled by any other test. It is also very clear that in combining tests greater importance or weight should be given to strong tests than to weak ones. By means of ingenious mathematical formulæ, such as the multiple-regression equation, there may be obtained the exact weights required to combine the various tests so as to make the best possible prediction or estimate of the aptitude in question. Also, by a simple manipulation of these formulæ there may be determined at the same time the correlation which the aptitude prediction made by means of the battery as a whole, when properly weighted, will have with the actual aptitude criterion. From this coefficient may be secured an excellent index of the prognostic success of the battery.

It often happens that under actual experimental conditions circumstances may be encountered which make it extremely difficult or even impossible for a particular worker to carry out in full detail all the steps of the process outlined above. In such cases there should be secured the best approximation to which the conditions permit. Sometimes the missing or imperfectly performed step may be carried out by a later worker who chances upon more favorable experimental conditions. It may be laid down as a general rule, however, that *none of the six steps may be left out without serious danger of loss to the prognostic efficiency of the resulting test battery.*

STEP ONE — ANALYSIS OF OCCUPATIONAL BEHAVIOR

With the general outline of test-battery construction before us, we may proceed to the detailed examination of the first step in the process — the psychological analysis of the behavior which the tests are designed to forecast. The purpose of this analysis is to identify as accurately as possible the various psychological traits or processes which are required for success. Naturally the elaborateness of this analysis will be determined to a considerable extent by the time available for the project as a whole. If there is only a short time available, the investigator may need to content himself with a relatively hasty and superficial analytical study of the occupational activity. On the other hand, owing to the inherent complexity of the determiners of human behavior, the most elaborate study is likely to result in an analysis which is by no means perfect. Up to the present time the psychological analyses of occupational behavior made in connection with aptitude-testing projects have been very largely of the hasty and superficial kind. Indeed, this phase of aptitude testing and of vocational psychology in general is in a highly embryonic state.

An example of such well-meaning but superficial and largely futile gestures purporting to be psychological analyses of occupations is given in the following list of qualifications supposedly necessary for a successful librarian: "A love of reading is not a major qualification, since a librarian has little time for reading. She needs accuracy, quickness, neatness, a pleasing appearance, and application necessary to do much routine work. All these qualifications must be founded on a sincere and active enthusiasm for library work."

It is possible that an analysis in such general terms may be better than none, though surely its value must be slight. Commenting on this particular analysis, Viteles remarks

without great exaggeration, "In such descriptions of mental qualifications the title of another occupation could, in many cases, be substituted for the one being described, and the statement would be equally accurate. Thus, for example, in the description for the mental qualifications for librarian, the term office worker, teacher, cashier in a department store, a scraper of celery in Campbell's soup factory, could be substituted for librarian, and the description would be equally applicable."

THE THREE PARTS OF A COMPLETED PSYCHOLOGICAL ANALYSIS

For purposes of aptitude testing the completed psychological analysis of an occupation or activity should consist, ideally, of three stages or parts. (I) The first part is an ordinary job analysis. This analysis is logical rather than psychological. It gives a detailed and rather minute list of the various part-activities of which the occupation is made up. (II) The second part of the analysis is a carefully controlled concrete study of a considerable number of individual workers, to discover in which part-activities of the occupation the more efficient individuals are chiefly superior to the less efficient. (III) The third part of the analysis is the extensive study of these critical part-activities found to be significant by part II. The purpose of this third part is to discover what traits of intelligence, capacity, temperament, will, etc., combine to produce efficiency in these particular segments of the behavior under investigation. The means by which the three steps of an occupational or aptitude analysis such as outlined above may be accomplished are various. In most cases a number of different methods may be employed jointly, the exact combinations depending upon the circumstances surrounding the particular investigation.

ANALYSIS FROM OBSERVING THE OCCUPATIONAL ACTIVITY

One of the best means of securing a preliminary view of the activity being analyzed is to observe attentively one individual worker after another as he goes through the various part-activities which constitute the occupation. If the job is the operation of an engine lathe, for example, the psychologist simply stands near the workman as he operates the lathe, observing as minutely as possible what he does. A very similar procedure would be followed in studying the behavior of a woman operating a hosiery-knitting machine. If the occupation were not stationary as was the case in the two examples cited, the psychologist will need to travel about with his subject. Thus Snow (77), in making the analysis of the job of Yellow Cab chauffeur, which extended over several weeks, spent some time "riding the meter" — i.e., riding in front with the driver. In this way he was able to observe at first hand the various emergencies which a chauffeur encounters.

A similar procedure would be followed in studying the job of a house-to-house salesman. In this latter case the method would have the defect that the presence of a third person might influence the behavior of both the salesman and the prospect. No doubt, also, the same holds to a certain extent of the behavior of all workmen who are being systematically observed. This is indicated rather clearly by the marked improvement in individual working efficiency usually noted under such conditions.

The observational method should be supplemented liberally, as opportunities permit, by careful inquiries of various persons who have a knowledge of the job. This will include talks with the workmen, with foremen, with managers, and with the instructors of apprentices in case the job is being taught systematically. A major obstacle ordinarily

encountered when seeking information in this way is that even thoroughly intelligent workmen and foremen experience considerable difficulty in understanding what is desired by a psychologist when they are questioned about the psychological aspects of their work. Their responses to the various work situations appear to be so much more manual than verbal that the assistance to be secured in this way is largely limited to the external aspects of the work. But even so, observation and judicious inquiry should place the investigator in possession of a fairly satisfactory analysis of the job into its various physical part-activities, which constitutes the first part of the occupational analysis. It should also result in numerous suggestions bearing on the second and third parts.

ANALYSIS FROM PERFORMING THE OCCUPATIONAL ACTIVITY

If the time available for investigation permits, the aptitude psychologist should also spend some time actually apprenticing himself to the job. This practice seems not only to have been advocated but followed by some of the most successful practical workers in the field. Thus Link (51), in analyzing the job of assembling gun parts in an arms-manufacturing establishment, was not content with merely observing the workers at a little distance and asking questions. Realizing the necessity for a first-hand knowledge, he "actually performed the various operations at a bench in the customary way, until he was able to do them with some celerity." Similarly Snow, in analyzing the job of chauffeur, actually drove a taxi. Viteles, in describing his method of analyzing the job of street-car motorman in Milwaukee, states (96): "The analysis in this case was based not only upon consultation with motormen, supervisors, instructors, and superintendents, but upon observation made in the course of a two-weeks apprenticeship as a motorman, during which time

the investigator ran street cars in every section of the city under a great variety of conditions."

"TIME STUDY" METHODS AS A MEANS OF ANALYSIS

A third method, the possibilities of which have not as yet been realized by investigators in the field of aptitude testing, is represented by the technique of the efficiency engineers such as F. W. Taylor (85). Taylor's chief contribution lies in what are known as *time studies*. He made elaborate studies of the time required by superior workmen to perform each of the part-activities making up various occupations. This was done largely for the purpose of making a scientific determination of piecework wage rates. Only a slight change in procedure would make the time-study technique extremely valuable in discovering which of the various part-activities of a given job will repay serious efforts at psychological analysis. The procedure would be simply to make careful studies of the time required by each of a number of superior workers to perform each of the various part-activities of a given occupation. These values would then be compared with similar time studies taken from the corresponding part-activities of a number of inferior workers. Those part-activities which in this way are found to show for a given quality of product little difference in speed between the two groups of subjects, could accordingly be dismissed from further consideration. It is obviously futile to attempt by tests to differentiate abilities where no difference in ability exists. The investigator would then be free to concentrate all of his energies upon the psychological analysis of those part-activities which show significant differences and where, consequently, an analysis may have at least a possibility of value in the choice of tests.

“MOTION STUDY” TECHNIQUE ADAPTED TO OCCUPATIONAL
ANALYSIS

A fourth method of some promise, particularly in the third and more psychological stage of the occupational analysis, also comes from the technique of the efficiency engineer. This is known as *motion study*. This particular phase of efficiency engineering has been associated largely with the name of Frank B. Gilbreth (24). Gilbreth's purpose in his motion studies was largely to find the “one best method” of performing a given piece of labor, usually of the repetitive type. To find this one best method he made most minute studies of the movements making up the separate part-activities of various jobs. Many of these were carried out by securing photographic records of the movements in question. By one method he would attach a small electric light to the hand of a manual worker. A stereoscopic camera was then exposed in such a way that the movements of the hand were recorded as a path of light on the two plates. These stereoscopic photographic records when properly mounted in a stereoscope showed the path of the light as standing out clearly in three dimensions. By having the current which supplied his electric light interrupted at a known constant rate by an electric tuning fork, the stereoscopic path of light became a series of arrow-shaped dots indicating not only the course of the hand movement but its direction and speed as well. If such minute analysis of behavior as this method renders possible could be made of the part-activities previously found significant by the time-study technique, it should evidently be a valuable supplement to the other available means of discovering the traits making for success in particular occupations.

EXPERIMENTAL ANALYSIS IN THE PSYCHOLOGICAL
LABORATORY

A fifth method by which the more psychological aspects of occupational analysis may be facilitated is by means of specific experimental attack. A typical example of such an attempt at experimental analysis is reported by Book (6). His problem was to determine whether or not a person's ability to master the art of typing depends to any important extent upon his "voluntary muscular control or general motor ability." Fortunately, he had available for experimental purposes a number of world's champion typists capable of typing between 137 and 147 words a minute, in addition to a large number of state and district champions in the various classifications. For purposes of comparison he had 65 ordinary students of typewriting from the School of Commerce and Finance of Indiana University. The experiment consisted of testing the speed of voluntary tapping with a telegraph key and comparing this record with the speed of typing.

Four different types of movement were measured for each arm and hand: (1) the rate per second at which these subjects could move their forefinger when the hand and arm were held in a definite and uniform way; (2) the rate at which they could move the hand, using the wrist as a hinge; (3) the rate and regularity with which they could move the forearm from the elbow joint, not moving the wrist; (4) the rate at which they could move the upper arm, using the muscles of the shoulder and upper arm.

It was found that the world's champions were quite uniformly superior to the lesser champions in their voluntary motor control, particularly in the elbow and shoulder movements. Probably one of the most significant results of the investigation for aptitude psychology was the observation that the world's champions showed a much greater superiority in the motor control of the left hand and arm than of the right hand

and arm. This investigation indicates rather definitely that the power of making rapid repeated movements with the hands and arms is intimately related to typing aptitude. Evidently a test somewhat like the tapping test used by Book would very probably be a useful unit in a battery of tests designed to forecast typing aptitude.

Finally it may be added that every aptitude-testing project carried through to completion as outlined at the beginning of the present chapter constitutes a kind of experimental analysis in much the same sense as the experiment of Book just described. All tests found in this way to correlate appreciably with a criterion may be taken as an assured starting point in subsequent testing projects involving that particular aptitude.

TYPICAL OCCUPATIONAL ANALYSIS — “BEAMER” IN TEXTILE INDUSTRY¹

As concrete illustrations of occupational analysis there are given in the following pages three examples from actual investigations. These examples purposely represent rather diverse aptitudes. No one of them is to be considered a perfect illustration of aptitude analysis. Each is intended to exemplify certain points not shown by the others. In order to relate the examples to the three parts of the analytical process as explained on page 286, the corresponding parts of the analyses will be numbered I, II, and III.

I

1. Run the new warp through the drum and up to the beam, by tying it to the old warp or by tying it to a leader.
2. Put lease rod on the warp by tying the gathered ends of warp to the rod.

¹ Beaming is an important occupation in the cotton-manufacturing industry. This is a typical factory job. The following is taken largely from Cades, Elliot (12).

3. Lay the warp on the raythe, getting the yarn placed properly according to the pattern card, counting all the yarn and splits (spaces in the raythe).
4. Put in the empty beam and tie the warp to the beam.
5. Adjust tension by use of weights.
6. Regulate speed by shifting belts.
7. When warp is finished, count and pick a lease.
8. Take the finished warp out of the frame ;
 - a. Get the right tie yarn.
 - b. Get the chain for cut marks.
 - c. Mark off the warp whenever the bell rings.
 - d. Tie up all breaks of yarn, making sure that the yarn is not crossed and letting no loose ends run through.
 - e. Oil and clean machine and drums.

II

There was apparently no attempt in this study to discover in which of the part-activities listed under (I) the superior workers excelled the inferior ones.

III

The psychological analysis in this case is notable in that it follows a rather elaborate and formal system. The analysis, when completed, takes the form of a *job psychograph*, as contrasted with an individual psychograph. A printed form such as that given below is designed to be used with all aptitudes alike. For each of the thirty traits listed a value of 1, 2, 3, 4, or 5 is assigned by the investigator at the conclusion of his analysis. The numbers represent degrees to which each specific mental ability is judged essential for success in the occupation. The significance of each number is as follows :

1. Negligible in amount.
2. Barely significant.
3. Significant.
4. Of great importance.
5. Of utmost importance.

JOB PSYCHOGRAPH OF BEAMER
(After Viteles and Cades, *op. cit.*)

1. Energy	—	×	—	—	—
2. Rate of discharge	×	—	—	—	—
3. Endurance	—	—	×	—	—
4. Control	—	×	—	—	—
5. Coördination A	—	—	—	×	—
6. Coördination B	—	—	—	×	—
7. Initiative	—	×	—	—	—
8. Concentration	—	—	—	×	—
9. Distribution (of attention)	—	—	×	—	—
10. Persistence	—	—	×	—	—
11. Alertness	—	—	×	—	—
12. Memory span	—	—	—	×	—
13. Visual discrimination	—	—	—	×	—
14. Auditory discrimination	×	—	—	—	—
15. Touch discrimination	—	—	×	—	—
16. Space discrimination	—	—	—	×	—
17. Form discrimination	—	×	—	—	—
18. Accuracy	—	—	—	—	×
19. Visual memory	—	—	×	—	—
20. Auditory memory	×	—	—	—	—
21. Kinæsthetic memory	×	—	—	—	—
22. Understanding	—	—	×	—	—
23. Observation	—	—	—	×	—
24. Planfulness	×	—	—	—	—
25. Intelligence	—	—	×	—	—
26. Intellect	—	×	—	—	—
27. Judgment	—	×	—	—	—
28. Logical analysis	×	—	—	—	—
29. Language ability	×	—	—	—	—
30. Executive ability	×	—	—	—	—

TYPICAL OCCUPATIONAL ANALYSIS — TAXICAB DRIVER¹

I

1. Drivers work six days a week in one of the following shifts:

5 A.M. to 3 P.M.

8 A.M. to 6 P.M.

3 P.M. to 1 A.M.

5 P.M. to 3 A.M.

Drivers must work Saturdays and Sundays.

2. Driver is assigned to a shift, reports to Yellow Cab garage nearest his home, changes to a complete Yellow Cab Company uniform.

3. Punches clock and reports to floorman (assistant to garage manager). Assigned to car. In case of old drivers, the same car is assigned to them each day.

4. Driver is given "meter sheet" which bears his name and payroll number. On it he must note in the proper columns the number of passengers, the time and place of picking them up and of discharging them, and the fare.

5. Driver calls at equipment window and gets his equipment — 3 skid chains, jack, and motor crank.

6. Driver must see that crank case is full of oil, gasoline tank full of gas, radiator full of water, and tires contain sufficient air.

7. If raining, driver must put chains on two rear tires.

8. Driver cranks motor, gets in driver's seat, tests brakes and lights, and leaves garage.

9. If driver runs out of gas or oil, he must report to nearest Yellow Cab garage to get new supply. The amount he gets is recorded on his meter sheet.

10. If motor breaks down, driver must go to the nearest Yellow Cab garage. If not possible to move he must tele-

¹ The following is taken largely from Snow (77).

phone for service. He is not permitted to remedy any motor difficulties.

11. Driver always carries spare tire, and in case of tire trouble, must change.

12. In case of accident, driver must fill out accident form supplied him by garage floorman.

13. When time to report back, driver goes to his garage and,

- (a) meter tape (on which the fares register) is removed,
- (b) crank case is filled with oil,
- (c) gasoline tank is filled, and quantity put in marked on meter sheet,
- (d) radiator is filled with water,
- (e) driver's equipment (crank, jack, and chains) checked in by driver or, if cab is to be sent right out again, he signs equipment over to driver who is to supplant him.
- (f) driver checks in the meter sheet, together with the money he has collected.

14. Driver can solicit business in any part of the city. He can do this in various ways :

- (a) He can report to nearest stand (branch office). Agent here may send him to fill a call that has come in to him.
- (b) He can station himself "in line" at door of some hotel, public building, dock, railway station, theater, or some important street intersection.
- (c) He can move along close to curb slowly, on the alert to pick up passengers.
- (d) If passenger takes him to outskirts of city, he can telephone traffic department for any orders it may have in that locality.

15. If engaged by a customer, driver must open door of cab, help passenger into cab, place any luggage in front of

cab, and ask the destination. (Driver must be familiar with streets and important buildings in city, and in addition carry a guidebook in which to look up any with which he is not familiar.)

16. When driver arrives at destination he must help passengers out of cab, help them with any luggage, and on getting fare, tender a receipt on which fare has been automatically registered.

17. Driver must be courteous. If necessary he must escort passengers to door, and give aid to cars in trouble. He must get his passengers to their destination as quickly and as safely as he can.

18. The drivers' earnings are on a common basis. They vary with his length of service and the brand of service which he has given. Previous to driving a cab must have passed through the employment department, the driving school, and the sales school.

II

No special investigation was carried out to locate the prognostically significant part-activities of this occupation. The nature of the occupation and the experience of the company, however, were such as to show with considerable clarity that failure or success as a driver consisted very largely in failure or success in avoiding accidents. Snow therefore concentrated his efforts at psychological analyses largely upon this particular phase of the work.

III

An analysis of the accidents of 3000 drivers over a period of six months revealed rather definitely that there was a definite group of accident men. It was found, for example, that 18 per cent of the drivers were responsible for 46 per cent of the accidents. Study was made both of the accident men and of the accidents. Drivers were questioned immediately after

their accidents, to get a line on their mental condition and their excuse for the accident. As a result of this intensive study of accidents the investigator came to the conclusion that accidents resulted from five more or less distinct psychological causes. At the same time the general nature of a test suitable for detecting the tendency to each of the five types of accidents was decided upon.

1. *Recklessness.* By this is meant a deliberate lack of care. Examples of recklessness are speeding, passing traffic lights, and running up the left side of the street.

It was concluded that for a test to be suitable for detecting this tendency in a prospective driver, it must put the subject in a situation in which there is both an urge for speed and at the same time an urge for care, as is the case in his work. Not only this, but the subject must not know that this particular part of his test behavior which is of interest to the psychologist is under observation. Otherwise any natural tendencies to recklessness might not appear at all. Therefore the test should involve the subject being directed in some detail to do one thing which incidentally involves doing a second thing which easily reveals recklessness.

2. *Carelessness.* An example of this is where the driver goes out with a defective car (e.g., a faulty steering apparatus), because he does not take the pains to inspect it properly before starting.

It was thought that for a test to detect this tendency it must present the subject with the task of performing several operations in the shortest time possible, yet without injuring the materials employed in the experiment. The materials and the general situation should be such that there would be an opportunity for either a response perfectly safe for the experimental materials or one which would involve great risk. The score would be based on the care and success of the subject in avoiding the risks.

3. *Emotional instability.* The inability to respond promptly and appropriately in moments of danger. At such times the emotionally unstable driver loses his presence of mind and there results a collision.

It was decided that to test this trait the subject should be given a task that would occupy his attention very completely, but with instructions that in case anything serious should go wrong with the apparatus he should instantly perform some particular emergency act. Then when S is busy performing his task, the experimenter arranges it so that the subject shall be stimulated simultaneously and unexpectedly by a loud explosion and an electric shock. The time and regularity of his emergency response should be taken, together with the emotional disturbances within the body of the subject as shown by suitable apparatus.

4. *Inability to foresee what the other person is going to do.* A general intelligence test of the pencil-and-paper variety was judged as most likely to detect the tendency to this type of accident.

5. *Miscellaneous physiological defects.* This is illustrated by the driver whose reaction time varies considerably from the average, who is easily distracted, has an abnormally short range of observation, or poor ability to estimate speed of objects or the distance between moving objects, or who has little muscular resistance to fatigue.

It was judged that a number of different tests would be necessary to detect the somewhat miscellaneous traits here assembled. As an example, it was thought that the ability of the subject to judge the speed and relative positions of moving objects might be tested by setting up an apparatus with objects moving at known speeds and instructing S to make rapid judgments regarding them.

TYPICAL OCCUPATIONAL ANALYSIS — FREE-HAND DRAWING

I

This aptitude, in contrast to the two examples of occupational behavior already considered, offers little opportunity for breaking up into physical and objective part-activities.

II

For the same reason there was no opportunity for seeking special prognostically significant part-activities.

III

It must first be observed that the aptitude at present under consideration is distinctly limited. It does not involve anything of the æsthetic. Beauty is not involved. Color does not enter. Neither do light and shade. The aptitude under investigation is merely the power to represent objects in perspective by means of black lines on white paper.

1. One of the most obvious processes observed when watching a person make a sketch is the fact that he first looks attentively at some part of the object and then works at the corresponding part of his sketch. This suggests that a capacity for remembering a large number of visually perceived details for at least a short time might be an important component of free-hand drawing aptitude.

2. In the technique of drawing it seems probable that in order to have a sense of proportion the student would need to be able to judge the actual size of objects regardless of variability of distance, at least within moderate range. This means that he should be able to judge accurately from the size of the visual image, in conjunction with his perception of distance, how large the object would look at a considerably different distance. This bit of analysis refers primarily to the perceiving or sensory receiving mechanism.

3. When the artist begins to draw a straight line it seems reasonable to expect that it will be of advantage for him to know with some precision where the line will, if continued, intersect other lines in the drawing. It also frequently happens that the combination of other lines associated with a straight line tends to make the straight line appear to be directed to a point at an appreciable distance from its true course. It therefore seems reasonable to expect that a test designed to measure the capacity of persons to tell where the lines in various surroundings would pass if extended, might have prognostic value.

4. In drawing activity various curved lines need to be extended just as noted above in the case of straight ones. Accordingly, a test involving the ability of a person in the completion of circles of various sizes and in various surroundings should be useful.

5. In perspective drawing the accurate reproduction of seen angles is extremely important. Accordingly, a test measuring the ability to reproduce angles should have prognostic value.

6. Since verbal intelligence (i.e., the general facility and accuracy in the use of words) has frequently been found associated with ability in the most varied kind of activities, such a test would be worth trying.

7. Because perseverance and patience (i.e., willingness to work at a sketch a long time in order to perfect it) may very likely contribute to ultimate success, a test designed to measure this trait should be worth a trial.

NOTE. The five remaining steps of the construction of the battery for free-hand drawing will be given in considerable detail in connection with the five following chapters, devoted to the respective stages of the process. The reader who wishes to study consecutively the entire procedure in deriving this battery may do so by reading pages 325 ff., 359 ff., 413 ff., 439 ff., and 477 ff.

CHAPTER TEN

THE ASSEMBLING OF A TRIAL BATTERY OF TESTS

HAVING performed the most adequate psychological analysis of the aptitude possible to achieve under the circumstances, our next task is to find tests which will measure the trait-complexes thus revealed. But, just as the task of analysis is difficult and the results nearly always far short of the ideal, so this task of fitting tests to aptitudes also is not easy and a considerable element of trial and error is involved. As a result, two or three times as many tests must usually be assembled for the preliminary test battery as are desired for the battery intended for final use. As a rule, between a half and two thirds of all tests are found upon preliminary trial to be for one reason or another without prognostic value, no matter how carefully the aptitude analysis and the test selections have been made. Accordingly, if a final battery of five tests is desired, between ten and fifteen tests should be assembled for the preliminary try-out. Such is our present ignorance of the subtleties of human aptitude.

INTRINSIC CHARACTERISTICS OF DESIRABLE TESTS

In selecting a test unit for an aptitude battery, many things must be considered. The main consideration, however, is that it shall correlate at least moderately with the criterion of aptitude efficiency.¹ Accordingly, the first consideration in the preliminary choice of a test is that it shall have a fair chance of showing such a correlation. The best possible indication of this is that the test has shown such a correlation in some previous investigation with the same or a very similar aptitude. *A careful piece of experimental work supporting*

¹ There is a technical exception to this. See pages 450 ff.

a presumption of correlation should outweigh almost any amount of negative presumption arising from even the most painstaking observational analysis.

A second principle is that it is usually wise to include one or two tests involving the facile manipulation and the discriminating use of words. This is particularly true where the activity involves any obvious symbolic elements such as are found in the various levels of thinking and common sense. Such tests are also usually worth trying in aptitudes largely motor, in case these are at all complex, even if there are no obvious verbal elements involved. Good tests for this purpose are verbal-analogies tests, mixed-relations tests, the matching of proverbs, synonym-antonym tests, completion tests, and many others (see pages 84 ff.).

A third principle favoring the choice of test units having high prognostic value is that we should avoid the ordinarily fatal error of attempting to test for highly generalized traits, such as: ability to observe in general, discrimination in general, quickness in general, industry in general, and so on. Very few human traits are sufficiently generalized for such attempts to succeed. On the contrary, an aptitude test should usually adhere as closely to the concrete activity of the aptitude as possible. For example, in assembling a battery designed to predict aptitude in free-hand drawing it would be unwise to construct a test calculated to measure capacity for observation in general, in which the subject would be required to *describe* the various things seen, heard, or otherwise observed under some standard conditions. For this aptitude the test response, instead of being verbal, should be *manual* because drawing is a manual activity. Not only this, but the test responses should be of the same general kind of manual activity as that found in the aptitude — i.e., making representative marks on paper with a pencil. As a rule, the more strictly motor an aptitude is, the more strictly the principle

holds. Indeed, the principle involved here is nothing more than that of the miniature test applied to the somewhat restricted trait aggregates isolated by the aptitude analysis.

In the selection of test units which shall have as high a correlation with the criterion as possible *it is equally important that at the same time the various units chosen shall correlate as low with each other as possible.*¹ To this end it is usually desirable while assembling the tests to have them, as regards the nature of the activity, as widely different from each other as possible while at the same time yielding the desired correlation with the criterion. This principle is extremely important though frequently violated. It will be taken up in detail in Chapter XIII, in connection with the choice of the test units to make up the final battery. The theoretical basis for it has been given in Chapter VIII.

ADMINISTRATIVE CHARACTERISTICS OF DESIRABLE TESTS

But the prognostic value, while of prime importance, is by no means the only factor to be taken account of in the choice of tests. Numerous other features, chiefly of an administrative nature, must be considered. One of these is whether the test requires such care in giving that only a highly trained psychologist is capable of administering it. Such a characteristic is a distinct though by no means fatal defect in a test. Other things equal, a test which may be given in a satisfactory manner by an ordinary person with no more than high-school training would be superior. Cases in point are the Binet-Simon Tests and tests involving the manipulation of delicate and complex apparatus. In giving the Stanford Revision of the Binet Tests properly, the examiner should know verbatim the prescribed verbal form of presenting each of the 83 different test items. The material to be memorized before these

¹ See pages 254-255, especially Table 52, and also Figures 34 and 35.

tests can be given properly amounts approximately to three thousand words. In addition, the examiner must know many special gestures and technical manipulations as well as a number of things which he must *not* do. By way of contrast may be considered the Terman Group Test of Mental Ability. This battery requires no rote memorizing whatever on the part of the examiner. He has little to do except to time the work intervals and occasionally read brief instructions.

A second administrative consideration of importance in the choice of a test unit is the time required of the examiner in giving it. It must never be forgotten that the basic value of a test unit is a function of the amount and importance of the prognostic information yielded by it, per unit cost. At bottom this is why it is undesirable for a test to require an expert to give it, as already pointed out. For the same reason it is desirable that the test should require, per subject tested, as little of the examiner's time as possible. Here again the Binet Tests may be contrasted with the Terman Group Test. By the former, from sixty to ninety minutes are required to test a single person. By the latter, forty minutes are sufficient to test a hundred or more persons, which averages less than half a minute of the examiner's time per subject. It may be added that the same principle of economy also holds, to a certain extent, of the time required of the subject.

A third administrative feature is the cost of scoring and otherwise manipulating the test results so that they may be available for forecasting purposes. If a test can be scored only by an expert, this obviously adds to the cost and constitutes a defect in the test. If the test requires a long time for scoring, this also is a defect. An example of a test suffering from these defects is the circle-completion test described on page 333. This test not only consumes considerable time in scoring but involves the use of a specialized mechanical

device requiring some skill to manipulate. There is, in addition, the cost of the mechanical device itself. By way of contrast may be mentioned Brigham's opposites test (page 332), which is highly efficient from this point of view. This test may be scored in a few seconds by placing a transparent key over the blanks at the right-hand margin and simply counting off the correct responses.

This does not mean that only tests capable of being scored in a summary fashion should be used. It means merely that tests should be scored in as economical a manner as the trait aggregate being measured will permit. If a test samples an important aspect of an aptitude not reached by any other test, it might very well be included in a battery even though it requires considerable time for scoring. Whether or not such a test should be included depends on whether or not the amount and importance of its contribution to the prognostic potency of the battery outweigh its cost.

THE UTILIZATION OF TESTS ALREADY AVAILABLE

In assembling the test units for the preliminary battery, there are two general procedures which may be followed. As often as not both are used in the same project. One procedure is to select appropriate tests from those already made available by the work of other investigators. Because of the obvious economies involved this source of "ready-made" tests should always be fully canvassed at the outset. The second alternative is to devise special tests to fit the peculiar psychology of the aptitude under investigation. As a compromise between the two procedures may be mentioned the modification or adaptation of already existing tests for special purposes.

By far the most convenient test units immediately available to the aptitude psychologist are those which have been published for commercial distribution. These tests are

usually of the pencil-and-paper variety, and the majority (though not all) of them are designed to test average scholastic aptitude and are referred to ordinarily as tests of "mental ability" or "general intelligence."¹

A second source of information concerning tests which have become more or less standard is found in books devoted to the subject of psychological tests. One of the most complete and valuable of these is Whipple's *Manual of Mental and Physical Tests* (Warwick & York, Inc.), which appears in two volumes. Other works giving considerable space to concrete descriptions of tests are Hollingworth's *Vocational Psychology* (D. C. Appleton & Co.), Link's *Employment Psychology* (The Macmillan Company), Kornhauser and Kingsbury's *Psychological Tests in Business* (University of Chicago Press), and Burt's *Principles of Employment Psychology* (Houghton Mifflin Company). Downey's *The Will-Temperament and Its Testing* (World Book Company) describes and gives references to a great variety of tests designed to measure traits of character and temperament other than abilities or capacities.

A third source of information concerning ready-made tests is found in certain journals devoted in part to applied psychology. Among these may be mentioned: *The Journal of Educational Psychology*, *The Journal of Applied Psychology*, *The Journal of Educational Research*, *The Journal of Personnel Research*, *Psychological Monographs*, *Industrial Psychology*, *British Journal of Psychology*, *The Journal of the National Institute of Industrial Psychology* (England).

¹ The leading publishers of such tests are:

World Book Company, Yonkers-on-Hudson, New York

C. H. Stoelting Company, Chicago, Illinois

Bureau of Publications, Teachers College, Columbia University, New York

Public School Publishing Company, Bloomington, Illinois

Bureau of Personnel Administration, Washington, D. C.

SOME ESPECIALLY DEvised TESTS USUALLY REQUIRED

As already stated, it is much more economical of time and energy when assembling an aptitude battery to use, so far as possible, test units which have already been worked out. It usually happens, however, that even after the most diligent search, no existing test can be found which appears likely to sample one or more of the test aggregates making up a given aptitude. In such cases special aptitude tests must be devised to fit the particular conditions. Fortunately there has emerged from the extensive experience of psychologists in test construction during the last few years a substantial body of knowledge regarding methods of constructing effective tests.

For the purposes of an exposition of test-construction methods, it will be convenient first to divide aptitude-test units into pencil-and-paper tests and apparatus tests. Pencil-and-paper tests may themselves be divided roughly into verbal tests and performance tests. Examples of verbal pencil-and-paper tests are Brigham's opposites test (page 332) and the various units of the Army Alpha battery. Examples of performance pencil-and-paper tests are Downey's slow-movement test (page 329), tapping on a paper as rapidly as possible, the line-extension test (page 336), and the circle-completion test (page 337). Examples of apparatus tests are: tapping with a telegraph key, Whipple's steadiness test, the coördination tests for the engine lathe (page 68), the peg-board test (page 75), the 3-hole test (page 106), etc.

PENCIL-AND-PAPER TESTS — ADVANTAGES

Of the three classes of tests, the verbal pencil-and-paper tests have been most extensively exploited during recent years. One great advantage of these is that they are usually group tests. This fact alone tends to make them highly

economical. If suitably designed, they also have a number of additional advantages of great importance. One of these is that the time required by the subject in recording his responses may be reduced to a smaller minimum than with any other test, thus reducing the time required by the examiner to give the test. Another advantage is that the time required for scoring the test may also be reduced to a minimum, thus further reducing costs. A third advantage is that if the test is suitably designed, disagreements among different persons scoring the same test are practically eliminated. This insures the test a high degree of objectivity.

These various advantages are made possible, in part, by the fact that each test unit is composed of discrete items, each of which requires of the subject an extremely simple response which is definitely right or wrong. Each test unit may contain from 20 or so to around 40 of these items.

While such items may be made up in the greatest variety of forms, they fall roughly into three groups: (1) the recall type, (2) the recognition type, and (3) the true-false type. A series of examples will make clear both the general nature of the respective types and at the same time the reasons for the various advantages mentioned above.

THE COMPLETION FORM OF TEST

The completion form of test contains one or more blank spaces scattered through a printed sentence, which are to be filled in by the subject in writing. The most convenient form is usually to have a blank space at the end of a short sentence for the insertion of a single word, as follows:

SAMPLE OF A SIMPLE COMPLETION TEST ¹

1. The Percheron is a kind of ____.
2. The color of sapphires is usually ____.
3. The tendon of Achilles is in the ____.
4. The number of a Kaffir's legs is ____.
5. The forward pass is used in playing ____.
6. The saber is a kind of ____.

The great advantage of this form of test is that there is very little chance of a subject guessing the right response, as may so often be done by the other methods.

This form of test item, while being fairly rapid in giving, suffers from the fact that a considerable variety of response may be made by the subjects and these responses have various degrees of correctness. This greatly complicates the scoring and correspondingly increases the cost. For example, on item 6 any one of the following words might be written in the blank space, none of which would be wholly without relevancy: *sword, scimitar, dagger, knife, weapon, implement of war*, and so on. Unless care is exercised, the phrasing of the sentences may easily aggravate this defect. Suppose, for example, No. 4 had been written:

A Kaffir has ____ legs.

The number of more or less relevant responses becomes enormously increased: fat, round, long, thin, angular, strong, bandy, brown, smooth, ugly, and so on almost indefinitely.

THE MULTIPLE-CHOICE TYPE OF TEST

The scoring defects of the recall type of test item are largely eliminated in the other two types. This is attained, however, at the sacrifice of allowing the subject opportunity to make many correct responses by mere guessing. This is shown by

¹ Adapted from Army Alpha.

the following examples of the recognition type of test item. For purposes of comparison the same material is used as on page 310.

SAMPLES OF MULTIPLE-CHOICE TEST ITEMS

1. The Percheron is a kind of goat, horse, cow, sheep.
2. Sapphires are usually blue, red, green, yellow.
3. The tendon of Achilles is in the heel, head, shoulders, abdomen.
4. The number of a Kaffir's legs is two, four, six, eight.
5. The forward pass is used in tennis, hockey, football, golf.
6. The saber is a kind of musket, sword, cannon, pistol.

In the above test the subject is instructed to underline that one of the four last words which completes the sentence correctly, as illustrated in the first two items. The amount of time required of the subject in the mechanics of the test is seen to be exceedingly slight. Since the words are so chosen that only one word is right and the others are clearly wrong, the problem of scoring is greatly simplified as compared with the completion type. Unfortunately an accompanying disadvantage is the fact that any given person would get a rather large but unknown number of items right by pure guessing. The law of chance would yield in the above case one correct response out of every four items attempted. Unfortunately this law holds reliably only if an infinite number of items is used. In actual practice an appreciable amount of error is introduced into multiple-choice test scores from this purely chance source.

A certain amount of difficulty in scoring this test results from the underlined words being scattered in a random fashion over the right half of the page. The responses can be assembled in a single column, which presents the greatest facility in scoring, by simply numbering the various alternative words and instructing the subject to write the number of

the word chosen on a short dotted line at the beginning of the sentence, thus :

-² The Percheron is a kind of (1) goat (2) horse (3) cow
 (4) sheep
 Sapphires are usually (1) blue (2) red (3) green (4) yellow
 The tendon of Achilles is in the (1) head (2) shoulder
 (3) abdomen (4) heel
 The number of a Kaffir's legs is (1) two (2) four (3) six
 (4) eight
 The forward pass is used in (1) tennis (2) hockey (3) foot-
 ball (4) golf
 The saber is a kind of (1) musket (2) sword (3) cannon
 (4) pistol

THE TRUE-FALSE TYPE OF TEST

The true-false type of test resembles in many ways the multiple-choice type. A series of sentences is given in which approximately half are true and half are false, the sequence being random. The subject must indicate in each case whether the statement is true or false. The choice is thus reduced to two. A series of items of this type follows :

The Percheron is a kind of horse.....	<u>True</u>	False
Sapphires are usually green.....	<u>True</u>	False
The tendon of Achilles is in the shoulder.....	True	<u>False</u>
A Kaffir has two legs.....	True	False
The forward pass is used in hockey.....	True	False
The saber is a kind of pistol.....	True	False

Sometimes the words *Right* and *Wrong* may be substituted for the words *True* and *False*. Another variant of this test is to replace the words *True* and *False* by a short dotted line placed at the left of each sentence, on which the subject is directed to place a plus sign if the sentence is correct and a minus sign (or a zero) if the sentence is incorrect.

RELATIVE EFFECTIVENESS OF RECALL, MULTIPLE-CHOICE,
AND TRUE-FALSE TESTS

Since much material suitable for pencil-and-paper testing may be presented about as readily by one of the above modes as another, the question has naturally risen as to what is the relative effectiveness of the various methods. A number of comparative investigations have been conducted with this end in view. One of the most important of these was performed by Toops (94). He constructed three 50-item tests, one each of the following: recall type, multiple-choice with five choices, and true-false. The subject matter in the items of the respective methods was identical, the only difference between the three series being in the mode of presentation. All three forms of the test were administered to 124 subjects, but in different orders for different squads of subjects in order to control practice effects. He found that the average reliability coefficients were as follows:

RECALL	MULTIPLE-CHOICE	TRUE-FALSE
.618	.556	.507

Exactly as might have been expected from the relative amounts of chance entering in the various tests, the recall was the highest and the true-false was the lowest. Fortunately Toops also took the time required by the various subjects to do each of the three forms. This averaged per item for the respective forms:

RECALL	MULTIPLE-CHOICE	TRUE-FALSE
6.9 sec.	5.6 sec.	3.6 sec.

These results show a progressive decrease in time required from the recall to the true-false type. This last is, of course, in favor of the true-false test. He then applied the Spearman-Brown formula to see how high a reliability coefficient the multiple-choice and the true-false tests would probably

yield if they were increased enough in length so that each would require the same time to perform as the recall form. He found that when this was done there was very little difference, the resulting reliability coefficients being :

RECALL	MULTIPLE-CHOICE	TRUE-FALSE
.618	.607	.664

A second investigation, reported by Ruch and Stoddard (70), was essentially of the same type except that the number of multiple-choice tests was increased to three, making five forms in all: recall, 5-choice, 3-choice, 2-choice, and true-false. All forms alike had 100 items, and the subject matter of the corresponding items in all the tests was the same. These authors found reliability coefficients for 50 items to be as follows :

RECALL	5-CHOICE	3-CHOICE	2-CHOICE	TRUE-FALSE
.811	.796	.598	.737	.555

In general there is a progressive decrease in reliability as the chance element is increased, exactly as found by Toops. There is one exception to this, however, in the case of the 2-choice test and the true-false, each of which superficially presents one chance in two of guessing correctly. Despite this apparent similarity as regards chance, these two tests differ considerably in reliability, the true-false being distinctly less effective. The true-false, on the other hand, required about 10 per cent less time to perform. When the reliability coefficients for all five forms were corrected by the Spearman-Brown formula so as to show the probable reliability if the tests were extended in length so as to consume the same time as required by the recall type for full 100 items, the various forms showed much less striking differences :

RECALL	5-CHOICE	3-CHOICE	2-CHOICE	TRUE-FALSE
.896	.901	.806	.902	.664

The two studies summarized above are typical of the experimental work devoted to this subject. Except for variations in the reliability coefficients evidently due to accidental factors in the experiments, there seems in general to be no very great difference in the reliability of various test forms, *provided the time-length is equal*. So far as these results may be generalized we may say that a 3-choice test should contain about 40 per cent more items, and a true-false test from 80 per cent to 90 per cent more items, than a recall test to secure the same degree of precision. To increase in this way the length of the tests naturally increases the size of the blanks used, the cost of printing and paper, and the time spent in scoring, over that for the same type of test with the original number of items. There is also the additional fact that the number of items suitable for tests of this kind is frequently limited, which makes the indefinite extension of tests of the above types a practical impossibility.

INFLUENCE OF "GUESSING" AND "NOT GUESSING"

If no special instructions are given subjects in the various multiple-choice tests, some individuals will attempt nearly all items whether they know anything about them or not, whereas others who are more cautious will not attempt any items which they do not feel rather sure about. This raises the question what the influence will be upon the value of a test, of giving specific instructions to guess at every item whether it is known or not, as compared with instructions not to guess at any item unless fairly certain about its answer. It is customary to accompany the instructions "not to guess" by the threat of a penalty for all choices incorrectly made.

An extensive investigation centering about this problem was carried out by Ruch and Degraff (71). The tests in this case were: recall, 7-choice, 5-choice, 3-choice, 2-choice, and true-false. The results show rather clearly that emphatic

instruction not to guess improves the reliability of the various tests, raising the average reliability coefficient from around .78 to .88. The improvement was greatest in the forms where the guessing element was greatest, the true-false form rising from .64 to .89.

Of course the real interest back of all this discussion of reliability is the question of *validity*. Reliability is ordinarily of interest only as it has an indirect bearing on validity. In the above investigation the recall type of test was taken as the criterion of validity, and the various multiple-choice forms were correlated with it to determine their validity. Curiously enough, almost no difference was found by this test of validity between the results of instructions to "guess" and "not to guess."

In connection with the matter of guessing, there has risen a difference of opinion and practice as to method of scoring errors and omitted items. A common formula for computing such scores is:

$$\text{Score} = \frac{\text{Rights} - \text{Wrongs}}{N - 1} \quad . \quad . \quad . \quad (12)$$

where N is the number of choices offered the subject at each item. If there are only two items, as in the true-false form, this becomes merely the subtraction of the wrong items from the right ones. The experimental evidence on the value of using this formula seems to be somewhat conflicting. Present indications are that when this formula is applied before computing the reliability coefficients, there is more likely to be a loss than a gain, but when applied in the case of validity coefficients, there is a slight tendency for a gain to result. The general question of the method of scoring errors is considered on pages 441 ff.

SIFTING OUT THE ITEMS COMPOSING PENCIL-AND-PAPER TESTS

In the devising and choice of test items the investigator has no alternative at the outset but to judge as well as he can the suitability of each for the purpose intended. But just as in the matter of the relative difficulty of test units, it must be said here also that the impressionistic judgment, even of the most experienced, is anything but reliable. Thus of several items, all of which appear to be of equal prognostic value, some may be many times as valuable as others. This means that in every test unit the items of which are selected merely on the basis of the investigator's judgment, there is almost certain to be a large amount of "dead timber" — items which have little or no prognostic value. It is estimated that out of every 40 test items evolved in the ordinary way, from 20 to 30 are not only of no value to the test unit but, since they blur the significance of the really valuable items, are a positive detriment to the test as a whole.

The separation of the prognostically significant items from the valueless ones may be done by a very simple objective method. The method presupposes that a considerable number of subjects previously have been tested by means of the test items. In addition it assumes that the aptitude of each of these subjects in the occupation or activity under investigation is known. These preliminary subjects are divided into three groups on the basis of their aptitude efficiency — a best third, a middle third, and a worst third. Then the number of persons who succeeded on a given test item in each of the three groups of subjects is ascertained. This shows, within the limits of sampling error, whether an item has the desired prognostic or selective value. A thoroughly good item should show a high percentage of successes in the first group of subjects, a smaller percentage of successes in the middle group, and relatively few successes in the worst

group. A non-selective item, on the other hand, will show no corresponding progressive difference in successes through the three groups of subjects. All items should be rejected which do not show distinct indications of prognostic power.

The procedure may be illustrated by a concrete example. The Thurstone Technical Information Test is a test unit containing 100 items of the multiple-choice variety. It was desired to sort out from this total those items which were especially valuable in forecasting the aptitude for learning to operate the engine lathe. A group of freshmen engineers at the University of Wisconsin who were taking shop practice involving the operation of the lathe were given this test. At the end of the course the skill of each man in the operation of the lathe was determined by a rather elaborate objective method. The 77 individuals composing this group were then divided into the three talent groups as described above, and the number of individuals in each talent group who successfully performed each item of the test was determined. The results of this procedure for the first 12 items are shown in Table 62.

TABLE 62

SHOWING THE NUMBER OF FAILURES MADE ON EACH OF THE FIRST TWELVE ITEMS OF THE THURSTONE TECHNICAL INFORMATION TEST BY EACH OF THE THREE TALENT DIVISIONS OF 77 FRESHMEN ENGINEERS (After Gleason)

The items marked with an asterisk were chosen to make up the final test unit for aptitude to learn the engine lathe.

APTITUDE DIVISIONS OF SUBJECTS	ITEM 1	ITEM 2	ITEM 3	ITEM 4	ITEM 5	ITEM 6*	ITEM 7	ITEM 8	ITEM 9	ITEM 10*	ITEM 11	ITEM 12*
Best third	0	0	0	4	19	2	0	0	0	0	0	0
Middle third	0	1	1	8	15	5	0	9	1	1	0	5
Poorest third	1	1	2	6	17	17	2	8	3	9	1	11

An examination of this table reveals that only three of the items show any considerable selective value. Of the entire 100 items only 30 were included in the unit finally organized as a part of a battery to forecast aptitude to learn the engine lathe.

It was desired in the above investigation to have unquestionable evidence of the greater prognostic power of a small selected list of items as contrasted with a comparatively long series of unselected items. Accordingly, the same Thurstone test was given to a second and comparable group of freshmen engineers for whom there happened to be available the same objective criterion of aptitude for learning the engine lathe as for the first group of subjects. From the analysis of the results with this second group of subjects, a still smaller group of only 25 items was selected. This second list naturally differed slightly from the original selection of 30, owing to chance sampling errors. The individuals of the *original* group of subjects were then scored on the bases (a) of the number of successes on the entire 100 items of the original test and (b) of the number of successes on the 25 items of the list selected on the basis of the performance of the second group of subjects. It was found that the score on the unselected list correlated .39 with the criterion, whereas the score on the small selected list gave a correlation of .43. Besides reducing the test to a fourth of its original size, with corresponding economy in printing and time of administration, the correlation coefficients indicate an actual increase in the prognostic efficiency (E) of the test, of over 16 per cent.

It may be added that this general method of selecting items has been used by Otis and others in the construction of such tests, with excellent results.

THE ARRANGEMENT OF TEST ITEMS

It is important to observe that in the construction of verbal pencil-and-paper tests there must be included items sufficiently easy that the least talented person likely to be tested by them will be able to obtain a certain degree of success. On the other hand there must also be included items sufficiently difficult, so that the most talented individual will not make a perfect score. The reason is obvious. The function of a test is to distinguish the different degrees of talent found among the individuals of a tested group. If there are no easy items in a test unit, perhaps 25 individuals out of a hundred may fail to make any correct responses whatever and consequently they will all alike receive a score of zero. The test accordingly cannot discriminate between the various degrees of talent possessed by these 25 individuals. If, on the other hand, there are no very difficult items, perhaps 25 individuals out of a hundred may show no failures, all alike receiving the maximum score. In this case, also, the test cannot possibly discriminate between the various degrees of talent possessed by these superior individuals. The items making up a test battery, unless it be purely a speed test, should accordingly range from the very simple to the very difficult.

By a simple extension of the above reasoning we arrive at the additional conclusion that, so far as possible, the various items making up any test unit should differ in difficulty by approximately equal amounts. This is a counsel of perfection, however. Ordinarily the chance sampling of the items, together with the variability in individual responses to any given item, serves to differentiate the various individuals fairly well if the two extremes of talent are provided for.

When organized into a test unit, the various items should be arranged accurately in the order of increasing difficulty.

By this is not meant the order of difficulty merely as judged by the impression of the investigator or the joint judgment of several other persons, however expert they may be regarded. The point is that the difficulty of such items simply cannot be determined with precision by impressionistic judgments. It is true that such judgments have some value, and if the conditions of the investigation are such that no more precise method can be employed, the investigator must use this method. The approved method, however, is first to arrange the items in what appears to inspection as the order of increasing difficulty. The tests are then given to a large number of persons and the number of persons passing each test determined, much as described above in connection with the evaluation of the various items for selectivity. Indeed, the one procedure is quite sufficient for both purposes. The items retained in the series are then arranged in such a way that each succeeding item is passed by a smaller and smaller number of persons. It is true that the order of increasing difficulty for one person will not be exactly the same as that for another, but the above arrangement reduces the difficulty from this source to a minimum. With the items arranged in this way, the person of limited capacity finds at once those items upon which he can demonstrate his powers without wasting his time and depleting his courage by struggling with items which offer no possibility of success.

PENCIL-AND-PAPER PERFORMANCE TESTS

There is no sharp division between verbal pencil-and-paper tests and pencil-and-paper tests of the performance type. Verbal *instructions* of some kind are required for both. But in the case of the performance tests, just as with most tests involving apparatus, the comprehension of the instructions is not intended to constitute a problem for the subject and the response of the subject does not involve verbal elements.

As an example of the pencil-and-paper performance test may be considered the well-known Downey slow-movement test. The subject is presented with a sheet of paper containing many loops of a dotted scroll (page 363). The subject is instructed to trace this dotted line with a pencil but *as slowly as he possibly can* yet keep his pencil moving continuously. Here the language element appears in the instructions, which must be very explicit; but the capacity to move with great slowness is probably not itself dependent upon language. Other examples of this type of test are shown on pages 364 to 373.

Pencil-and-paper performance tests share many of the advantages enjoyed by the verbal type. Perhaps the most important single advantage is that they lend themselves readily to group testing methods. Secondly, as a rule the school training of most subjects has made them familiar with the materials, which are cheap and readily accessible. In addition it is probable that the variety of capacities possible to sample in this way is far greater than is generally supposed, though as yet these possibilities have been very little explored.

On the side of limitations it may be said that while some pencil-and-paper performance tests are as easily scored as any verbal test, others require both time and skill. The circle-completion test to be described later (page 370) is a case in point. When compared with the variety of behavior possible to sample by the use of apparatus, it is also seen that pencil-and-paper tests offer distinctly less variety, being limited largely to eye-hand coördinations of various kinds.

PROBLEMS CONNECTED WITH THE USE OF APPARATUS

Owing to the various limitations of pencil-and-paper tests, it often happens that one or more of the factors shown by the preliminary analysis as probably contributing to success on any given aptitude are not sampled by them. In this case the investigator must turn to apparatus. In general, apparatus should be used only as a last resort. It is costly and time-consuming to construct. It is likely to require considerable time to administer, since such tests ordinarily must be administered to individuals and cannot be used with groups. As a general thing, however, they have the advantage of resembling more or less some part of the aptitude or vocation under investigation. This is likely to be very convincing to subjects being tested. It gives them confidence in the value of the test and in this way may assist materially in securing their coöperation, which is indispensable.

As a rule, standard ready-made apparatus, or apparatus already designed and used by other investigators for somewhat similar purposes and found valuable, should be used where possible. The reason, of course, is the obvious economy involved. Frequently a modification of a standard device also may save the construction of a special piece of apparatus for the particular investigation.¹

But after all possibilities have been exhausted in utilizing the apparatus already available, there are likely to remain certain part-activities peculiar to the aptitude under investigation for which no test is provided. More frequently still it may be desired to construct a miniature test duplicating the essentials of either the whole or a large part of the aptitude under investigation. This generally means the construction

¹ In this country practically all standard psychological and testing apparatus may be secured from C. H. Stoelting Company, of Chicago.

of a more or less elaborate apparatus which is usually of unique design.

The knack of designing clever apparatus is probably a highly specialized ability, and no very definite rules may be laid down regarding it. It is possible, however, to state some of the characteristics of a good miniature apparatus. It should strike a happy mean between being so similar to the actual job that experience with the job will greatly influence the score, and being so different from the job that it will not sample in an adequate manner the behavior of the vocation. Secondly, the apparatus should be made as light and portable as possible. Often a simple, inexpensive device will be just as effective as a ponderous and expensive machine.

SELF-RECORDING AND SELF-SCORING APPARATUS TESTS

One very desirable characteristic of any aptitude apparatus is that the score should be recorded automatically by the instrument, so that the moment the test is over the examiner may glance at a dial and instantly read the score. Very often tests may be so designed that the activity of the subject will be of a repetitive character. The number of repetitive acts may easily be counted automatically either by an electric counter or, still better for most purposes, by inexpensive mechanical counters. A great variety of these devices appropriate for counting the number of revolutions or oscillations of various types are manufactured by the Veeder Company.

In cases where the behavior cannot be summarized mechanically in such a simple manner as suggested above, it will be desirable to have it recorded automatically in some way. If the energy involved in the behavior being measured is very slight, as in the case of the psycho-galvanic reflex, it may be necessary to use a beam of light, the oscillations of which

are photographed on a moving film (page 77). Less feeble processes may be recorded by a light wooden or metal lever

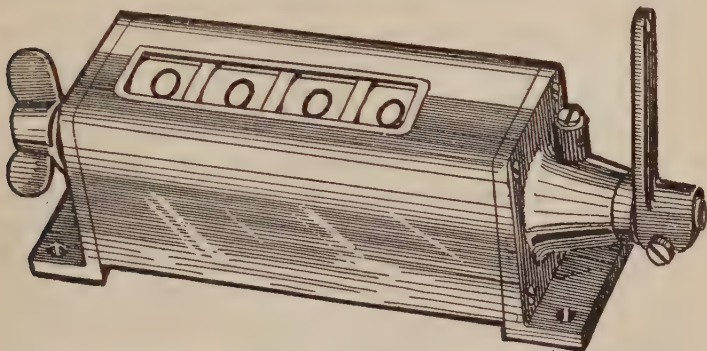


FIG. 37. A mechanical counter made by the Veeder Manufacturing Company, Hartford, Connecticut.

tracing a path by moving a point over a glossy paper lightly coated with smoke (page 79). Both these methods are slow and expensive. For most apparatus, especially of a miniature type, a graphic record of the subject's behavior may be secured either by direct graphic tracing on paper or by some indirect method involving electricity and electromagnets. If a long series of reactions is to be recorded, this may often be done very nicely by having electromagnets trace on adding-machine or other paper tape, by means of attached lead pencil or pen points, the various reactions in their exact temporal sequence. The paper may be moved beneath the magnets at a constant rate by rollers operated by a small electric motor.

DETAILED PROCEDURE IN ASSEMBLING A TYPICAL TEST
BATTERY — FREE-HAND DRAWING

By way of a concrete illustration of the methods of assembling a trial battery of tests there will now be presented the

various steps of actual procedure in the case of one of the aptitudes analyzed at the close of Chapter IX — that of free-hand drawing.¹ Since the aptitude activity is obviously itself of the pencil-and-paper variety, it was decided at once that pencil-and-paper tests offered with this aptitude an unusual prospect of success. Accordingly there were assembled in a booklet eight pencil-and-paper tests which could be given to groups of 50 or more subjects at one time. A description of the various tests, the reasons which led to their particular choice or construction, the method of administration, timing, etc., will be given in connection with each test. These details are presented in order to give the reader inexperienced in aptitude testing as vivid and realistic a view as possible of an actual procedure.

Test for Memory of Design

In the preliminary analysis of representative drawing, it appeared probable that this aptitude might be dependent in part upon the degree of ability to remember the appearance of an object being drawn, long enough to reproduce it with a pencil. Since no test of this particular ability was known, it was decided to devise a special test for the purpose. It was very evident at the outset that the presentation for this purpose of an ordinary object to be drawn would be useless, because an objective method of scoring the subject's response would be impossible. Moreover, the response would prob-

¹ This project is the result of the efforts of four individuals: Caroline Woods, Ula Strader, Selmar Larson, and the writer. Miss Woods carried out a preliminary investigation in which she tried out a number of tests. The battery which she finally evolved had a correlation yield of only .36. While her battery was of no particular immediate value, the relations found to exist between certain of her tests and the aptitude were of considerable assistance in making the preliminary choice of tests in Miss Strader's study here reported. The statistical work on Miss Strader's data was performed by Mr. Selmer Larson. All the work has been done under the direction of the writer.

ably reflect acquired skill in graphic representation quite as much as any memory function. These difficulties were finally avoided by presenting the subject with a design drawn on white cardboard, made up of simple straight lines (Fig. 38). In reproducing this design the amount of drawing skill required was reduced practically to the vanishing point. At the same time the problem of scoring was solved in a practical way by taking as the measure of success the number of the straight lines correctly reproduced. On the principle that the simpler parts of any test should come first (page 320), the designs are arranged in the order of increasing difficulty from the top downward.

For presentation, these designs were drawn in India ink on a piece of white cardboard 15" by 22" in size. With this design ready for exposure, the experimenter stood before the group of subjects and instructed them as follows:

"Pencils up! Turn over the first page of your booklets for the design test. On the other side of the card which I am holding are four groups of figures. As soon as I turn it over, *study every detail of each design very carefully* until I say 'Time.' Then I shall turn the cardboard down and you will make smaller drawings in your booklet of each of these designs. Be careful to draw them just as they are here. Remember: You study the design for thirty seconds but make no marks whatever. Then I shall say 'Time' and you will draw them on paper just as exactly as you can in all details. Do not worry about the shading. It is the shape that counts."

Directions Test

Since there is involved in most complex activity a certain element of symbolic behavior of a more or less verbal nature, it is nearly always well to include in a preliminary battery one or two tests dependent upon this group of determiners. In the present project two such tests were tried — the Woodworth-Wells directions test now to be described and

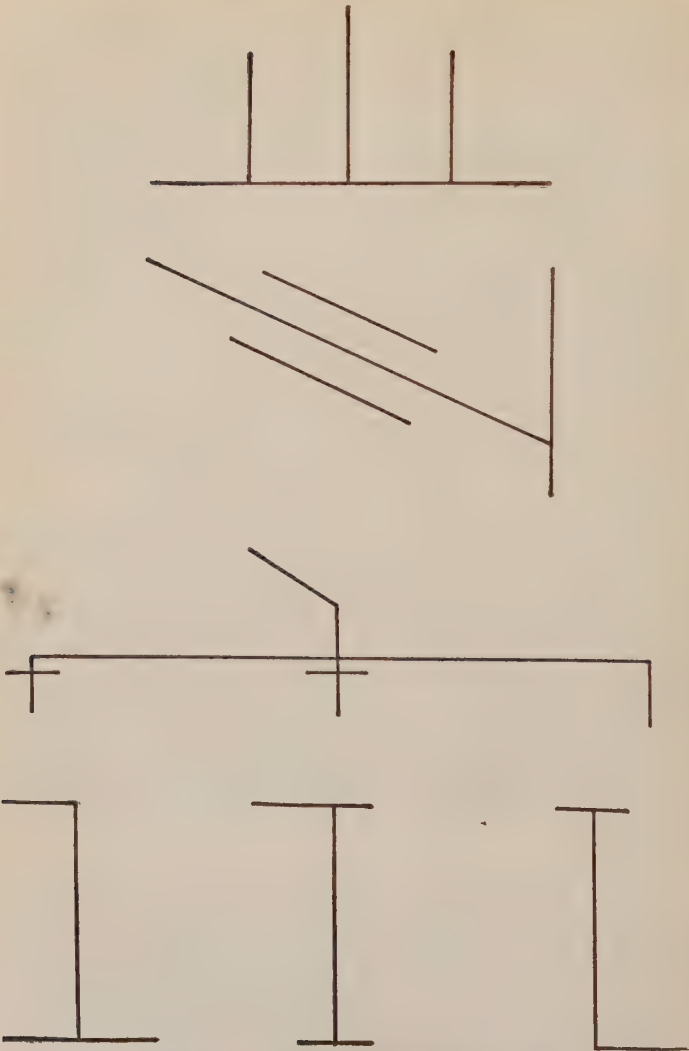


FIG. 38. Stimulus, memory for design test.

the Brigham opposites test to be described later. The Woodworth-Wells test (Fig. 46) is a very inexpensive test¹ and requires only one minute of actual working time by the subject.

The directions for the test were as follows :

“Pencils up! The next is the directions test. When I say go (but not before) turn over to the next test and do everything it tells you to, as rapidly and as accurately as possible. You can do all the things with your pencil. Remember: As rapidly and as accurately as possible. All ready, go!”

Downey's Slow-Movement Test

Since the analysis of drawing activity disclosed the need for considerable patience and the probability of a certain amount of inhibition of a tendency to hasty pencil work, at least in the early stages of learning representative drawing, it was thought possible that the Downey slow-movement test might make a useful member of the battery. A second feature strongly recommending this test was that it was so different from ordinary tests that in case it should show an appreciable correlation with the aptitude, it would not run much chance of being lost from duplication by some of the other tests. Since this test is published along with a number of others in a booklet,² it was necessary to buy a complete booklet for each subject, and cut out with scissors the scrolls used in the present test, discarding the remainder of each booklet. The piece of paper bearing the two scrolls (Fig. 47) was pasted to a sheet of plain typewriter paper of the size of our own booklet.

The verbal directions for the slow-movement test were as follows :

¹ Published by C. H. Stoelting Company, Chicago.

² World Book Company, Yonkers-on-Hudson, New York.

"Turn over to the slow-movement test. Now listen: So far, you have been going as fast as possible. This test is exactly opposite. Here you are to go as *slow* as you can. Look at the scrolls. When I say go (but not before) put your pencil at the beginning of the first scroll and begin moving along the dotted line just as slowly as you possibly can and still keep moving. The smaller the distance covered in three minutes, the better the score. If you get through one line before time is called, start in on the next. All ready, go!"

Reproduction-of-Angle Test

Since representative drawing frequently requires the accurate reproduction of observed angles, it was decided to include a test designed to sample this ability. Accordingly five angles (Fig. 39) were presented, one at a time, and the subjects were requested to reproduce the angles on a page of their booklet. Each angle was drawn in India ink on a separate piece of white cardboard 7" square. Three quarters of a minute was allowed to study and draw each angle.

The oral directions were as follows:

"Turn over to the angle test. On these cardboards I have five angles. I will show them one at a time and you will have three quarters of a minute to study and draw each angle. You may make the legs of the angle shorter, but make the angle itself exactly the same size as it is here. Draw the angle in the same position as shown, and do not move your paper around. All ready (holding up the first cardboard), draw angle 1."

After three quarters of a minute the cardboard was laid face downward on a table and the second angle was held up, the examiner saying:

"All ready, draw angle 2," and so on.

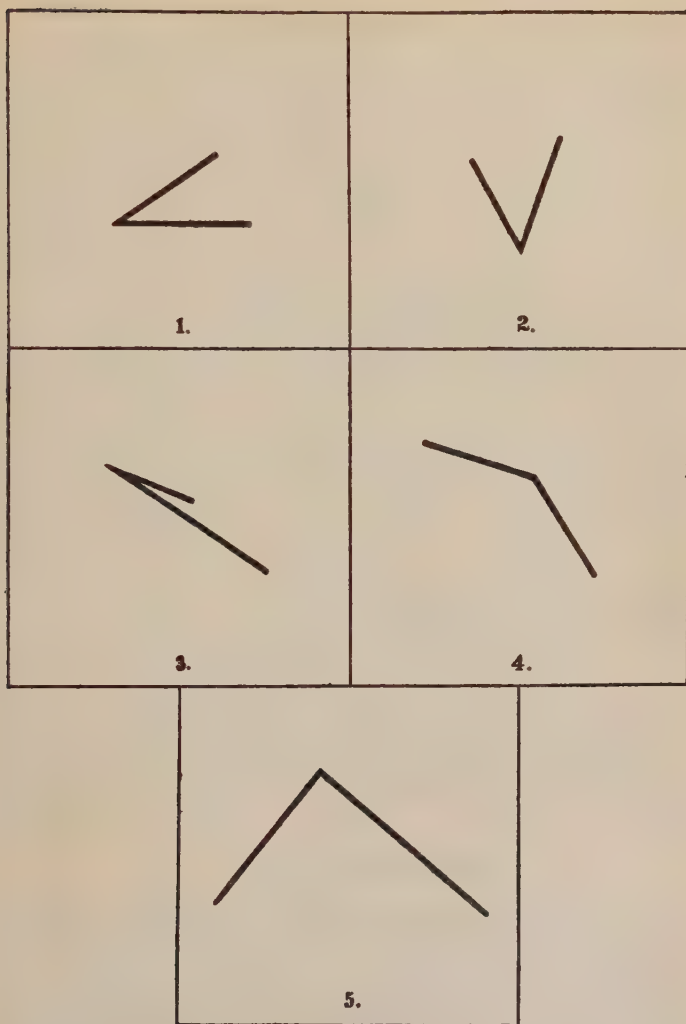


FIG. 39. Showing the angles presented to the subjects in the test for the reproduction of angles.

Opposites Test

As pointed out above in connection with the directions test, it was deemed desirable to include in the battery a test depending on the power to manipulate verbal symbols. A test of this kind, but differing rather strikingly in the superficial appearance of the activity required, is the Brigham opposites test (Fig. 49). It was hoped that the two tests might be sufficiently different so that both might be retained in the final battery in case they should both turn out to correlate well with the criterion. But at the worst, in case they should correlate too high with each other for both to be retained, we should at least be able to choose the better of the two for the final battery. This test was found in a booklet of tests entitled, *Psychological Examination for High School Graduates and College Freshmen*, organized under the direction of L. L. Thurstone. A full set of these booklets was bought, just as in the case of the Downey slow-movement test, and the page containing this test removed and inserted in the booklet of the present test.

The oral directions for the Brigham opposites test were :

“Pencils up! When I say go (but not before) you will turn over to the opposites test. The instructions are given at the top of the page. First be sure you understand what you are to do. Then do the test as accurately and as rapidly as possible. You will be allowed *five minutes*. All ready, go!”

Judgment-of-Size Test

Since correct representative drawing constantly involves the question of size as related to proportion, it was decided to introduce a test involving this behavior complex. It was reasoned that in drawing, the subject must make a reaction based jointly upon size and distance. No test of this kind was already available; so one was constructed. The four

figures shown in Figure 40 were drawn on separate pieces of white cardboard $8'' \times 10''$ and presented, one at a time. They were held up just above the experimenter's head and moved about slightly so as to face different parts of the room in order to prevent any bad angle of reflection from showing continuously in any one direction.

The oral directions for the judgment-of-size test were as follows :

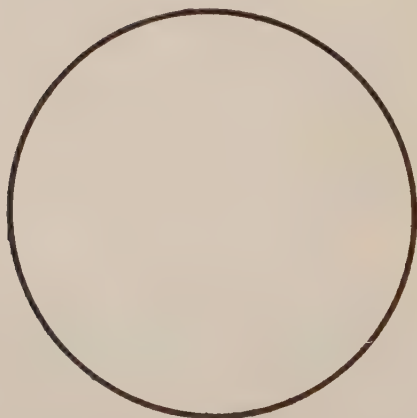
“Pencils up! I want you to draw these figures that I am going to show you, *in outline*, just exactly the same size as they are here. There are four figures. I shall show them to you one at a time. You will have about three quarters of a minute on each one. Draw each figure on its own sheet of paper. Now remember: Draw each figure outline exactly the same size it is here on this cardboard and in the same proportion. All ready: Turn over to the blank sheet for the square.”

It was found by trial that the circle and the cross required more than the three-quarters minute allowed; so the circle was given a full minute and the cross 1.5 minutes. Just before withdrawing one figure and presenting the next, the experimenter would say :

“All ready: Turn over to the blank sheet for the circle,” and so on.

Circle-Completion Test

As a second attempt to tap the complex often spoken of as the “sense of proportion,” the circle-completion test was designed. It consisted of five arcs of circles, all of different sizes, scattered over a page (Fig. 41, on page 337). The subject was asked to complete these circles as perfectly as possible.

Fig. 40 *a*.

FIGS. 40 *a* and 40 *b*. Showing the stimulus figures in the judgment-of-size test.

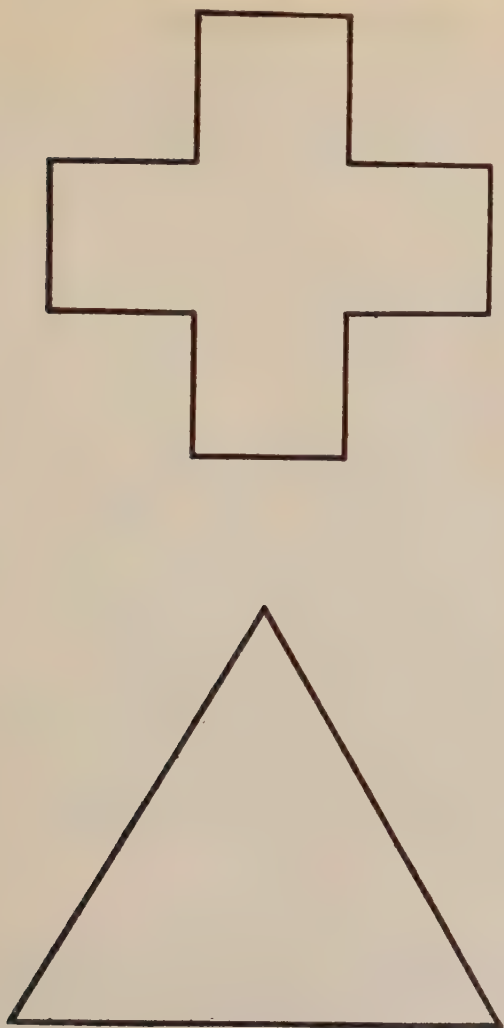


Fig. 40 b.

The oral instructions were as follows :

“Turn over to the circle-completion test. In this experiment you will be allowed one-half minute to complete each circle. When the signal is given, trace with your pencil the rest of the circle, making it complete. Do not start on circle 2 until you are told. Remember: Make the circles as perfect as possible. All ready, start on circle 1.”

While an apparent time limit was employed on this test, it practically amounted to a power test, since the time intervals were not rigid and practically all were permitted to finish. About three quarters of a minute was allowed for the first three smaller circles. The larger ones required longer, the largest one being given one and three-quarters minutes. When practically all subjects had completed a given circle, the experimenter would say: “Now start on circle 2,” or “Now start on circle 3,” and so on.

It is well for the reader inexperienced in the devising of tests to observe that this fairly complicated test was printed for use by an ordinary mimeograph, as was the line-extension test next to be described. The cost in each case was practically nothing, except the trifling labor of preparing the stencil. If a test so produced should later prove to be of value, a zinc etching would be made and the printing done on an ordinary press. Unfortunately, the changed appearance of the test in the latter case would probably be great enough to make desirable the establishment of a new set of norms.

Line-Extension Test

In free-hand drawing, when a person starts to draw a line, he ought to know where it will terminate. In cases where the lines of a large object are temporarily obscured by a small intervening object, the artist must be able to continue the line in its true position, despite the interruption. With

CIRCLE-COMPLETION TEST

In this experiment you will be allowed $\frac{1}{2}$ minute to complete each circle. When the signal is given, trace with your pencil the rest of circle, making it complete. Do not start on circle 2 until you are told. Make the circles as perfect as possible.

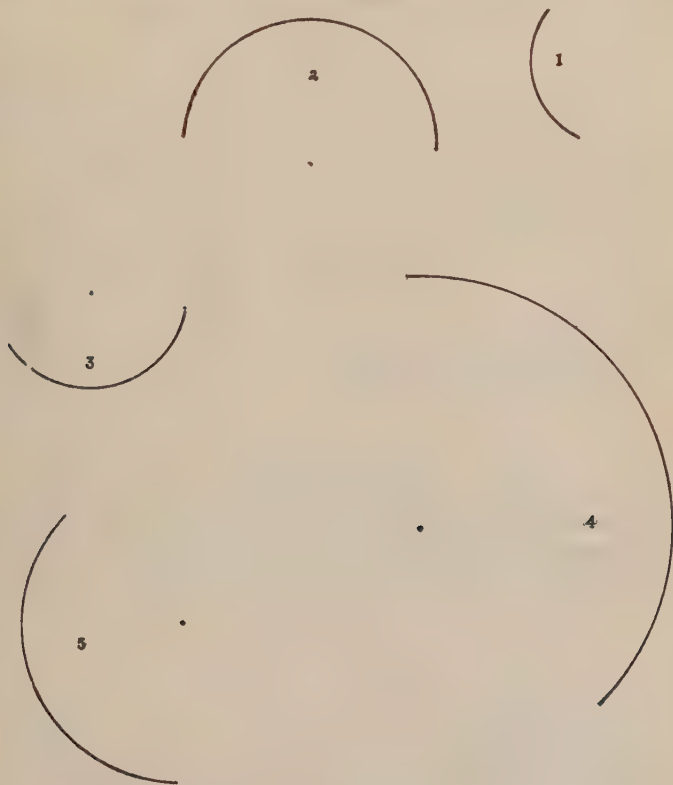


FIG. 41. Showing the form for the circle-completion test. (Reduced.)

LINE-EXTENSION TEST

First take your pencil and make a dot on the vertical line A' where horizontal line A would intersect if it were to be extended. Then do the same with the lines of B , C , D , and E . Remember, do not draw any lines or use any straightedge or sight along the lines. Locate the dots as accurately as possible.

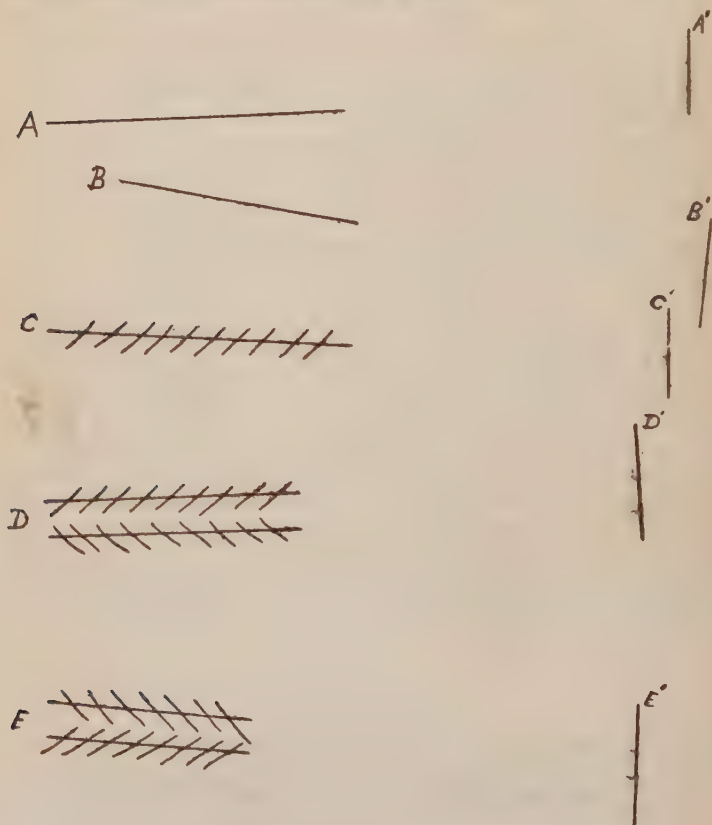


FIG. 42. Showing the form for the line-extension test. (Reduced.)

the hope of testing this ability the line-extension test was devised. Five lines or pairs of lines, varying from two to three inches in length, were drawn on the left side of a sheet of paper (Fig. 42). On the right side of the paper, between three and four inches distant, were drawn lines approximately an inch in length perpendicular to the direction of the first ones and so placed that if the horizontal lines should be extended, they would obviously intersect one or another of the perpendicular lines. As an added feature, several of the horizontal lines are cross-hatched so as to produce a well-known illusion of direction of extension. These cross-hatchings were inserted on the assumption that optical illusions distorting more or less the apparent direction of lines in objects are frequently encountered by artists. It was further assumed that for the drawing to be true the artist should not be misled by such illusions. Two minutes were allowed for completing this test.

The oral directions for the line-extension test were as follows:

“Turn over to the line-extension test. First take your pencil and make a dot on the vertical line A' where the horizontal line A *would* intersect it if it were extended. Then do the same with the lines B, C, D, and E. Remember: Do not draw any lines or use any straightedge or sight along the lines. Locate the dots as accurately as possible.”

CHAPTER ELEVEN

ADMINISTERING THE PRELIMINARY TEST BATTERY TO A TRIAL GROUP OF SUBJECTS

WITH the psychological analysis of the occupation completed and a preliminary battery of from eight to fifteen or more test units assembled, we proceed at once to *test the tests*. Experience has shown that if half of the test units survive the testing or validating process by showing that when incorporated into a battery they will contribute substantially to its prognostic potency, we shall be fortunate. Indeed, it is not unusual for as few as a fourth of the test units, or even less, to prove of practical value. All tests which do not thus demonstrate their worthiness of a place in the final battery must be discarded ruthlessly, no matter how promising they may appear in other respects. At the best, the presence of these inert elements in test batteries involves a waste of time and energy to administer and score them. At the worst, in case the various units of the battery are not scientifically weighted, these worthless elements may be distinctly harmful. They may actually destroy in considerable part the prognostic possibilities of the other really valuable test units. Unfortunately, the rigorous sifting of the useful test units from the worthless has not always been the practice among the makers of tests. Even at the present time psychological tests and test batteries are occasionally put forward, and even placed on the market, without this fundamental procedure. The present chapter gives the first step in "testing the tests" — the administration of the preliminary battery to a trial group of subjects.

AN ADEQUATE CRITERION SCORE MUST BE PROCURABLE

The securing of a suitable group upon which to try out a preliminary battery is the first and sometimes the most

serious problem encountered by the aptitude psychologist. Indeed, if such a group is not available, the testing project may as well be abandoned so far as any scientific possibilities are concerned.

The first consideration regarding the trial group is that the actual aptitudes of the individuals shall be susceptible of fairly accurate measurement. This measure is called the *criterion score*. Indeed, unless there is reason to believe that an adequate criterion score will be obtainable from any given group of subjects, no time should be wasted on them. Unfortunately, the securing of an adequate criterion score is often extremely difficult. This is one of the chief reasons that it is so hard to secure a satisfactory trial group. And even where an adequate criterion measure is really obtainable, the procedure by which it is secured is likely to be fairly technical. Indeed, the next chapter (XII) is devoted largely to a description of various experimental and statistical methods by which the numerous difficulties connected with obtaining the criterion score may be met.

NUMBER OF SUBJECTS NECESSARY FOR THE TRIAL GROUP

The second requirement of a trial group of subjects is that it shall be of considerable size. A trial group of from 20 to 25 individuals, for example, is entirely too small to yield results worth the trouble involved in securing them. Probably 50 or 60 subjects are about as few as it is profitable to work with. Generally speaking, as many subjects should be secured as possible. Satisfactory results may be obtained with groups ranging from 100 up to several hundred individuals.

As a rule, even where the conditions for the securing of a criterion score are most favorable, a certain per cent of the subjects will fail to be present when the tests of the battery

are given, because of illness and a great variety of other reasons. These subjects should usually be excluded from the investigation. Of those who come through with a complete set of test scores an additional per cent will fail to complete the training by which the potential aptitude possessed by them may become actual. These are also lost to the investigation. And of those who complete the training a certain additional number are almost sure to fail for one reason or another to receive trustworthy scores on the all-important criterion. The total of these various losses is such that the group with which an investigator begins, even under favorable circumstances, should be from 25 per cent to 50 per cent larger than the minimum with which he is willing to complete the project. As a concrete example of this difficulty it may be mentioned that in the investigation of free-hand-drawing aptitude, out of about 75 subjects apparently available at the outset, only 48 were actually available for final utilization in evaluating the tests.

THE TRIAL GROUP SHOULD BE SIMILAR TO THE GROUP
FOR WHICH TESTS ARE INTENDED

A third requirement is that the trial group shall be, as regards previous training, experience, and aptitude, a random sampling of the population which the final battery is designed to test. Previous training and experience should be similar in both populations because there is always considerable danger that previous learning and skill may influence certain of the test scores, quite apart from any special aptitude possessed by the subjects. Such effects disturb the correlation coefficients, which in turn distort the weights later assigned to the various test units of the battery finally selected. A distortion of the true weights used in making predictions naturally reduces in turn the prognostic potency of the test battery. In case the trial group is of noticeably

different average aptitude from the subjects upon whom the final battery will be used, there is also danger of setting up misleading norms. Moreover, in case the trial group is more variable the prognostic value of the tests is likely to appear greater than it really is, while if it is less variable, the prognostic value is likely to appear less than it really is.

Unfortunately, the requirement that the trial group shall be strictly comparable to the population upon whom the tests will later be used must frequently be violated at the beginning of an investigation, though sooner or later the condition must be conformed to if the tests are to possess their maximum value. The reason for this frequent necessity of disregarding the requirement is that it often consumes a very long time for individuals to go through the training necessary to reveal their actual aptitudes in particular lines. This period may vary from a few months to many years. In order to avoid this difficulty psychologists frequently are forced to use as a preliminary group subjects who are, at the time of testing, in practically their final stage of training. Where such a group is employed, the various distortions mentioned above as resulting from the training process and the selective process usually accompanying it should always be kept in mind. In such cases the resulting battery should be regarded merely as a preliminary approximation to the true form suitable for forecasting the aptitude of genuine novices.

TEST UNITS MUST BE IN THE FORM INTENDED FOR THE
FINAL BATTERY

Before the actual administration of the battery to the trial group of subjects, the tests should be in as nearly their perfect and final form as possible. Apparatus, for example, if any is used, should be tried out carefully to make sure that it is in perfect working order. The instructions to the sub-

jects should have been precisely formulated and the whole tried out on a number of typical individuals of the sort later to be tested. These individuals may be spoken of as the *pre-trial* subjects. Often gross and fatal defects in the clarity of the instructions or the adaptability of the apparatus for the purpose designed may thus be discovered and corrected before the beginning of the testing on the regular trial group. Almost never can a test be designed that will not require some such adjustments. This preliminary trial is of the greatest importance because, if a defect should be discovered only after a considerable part of the precious trial group had been tested, the investigator will be faced with a choice between two alternatives both of which are likely to spell experimental failure. If he rectifies the technique, the scores of those subjects taking the test under the two different techniques are not comparable and the test can therefore not be used. Of course, all those subjects tested by the defective technique may be excluded from the experiment. But this will reduce the effective numbers of the trial group, which will impair the reliability of the final results, to say nothing of the waste of time and energy involved in the testing of the excluded subjects. A second alternative is to continue the defective technique throughout the remainder of the trial group. If this is done, the test in the defective form may be worthless. If, on the other hand, it should turn out that the test, even in its defective form, has some value, and it should be desired to include it in the final battery but in an improved form, the norms of performance and the weight received by the particular test from the trial group will not correspond to its new form, since it will practically have become a new test. It is difficult to overemphasize the desirability of having the tests in their final stable form when the testing of the trial group of subjects begins.

THE ORDER OF TESTS IN THE BATTERY

The order in which the tests are given is a matter of some importance. The problem of arrangement is complicated by the fact that, at best, several of the tests in the preliminary battery will not find a place in the final battery. As far as possible, however, the tests should be given in the same order as in the final battery, regardless of which tests ultimately find a place there. In deciding upon the order, two general principles should be observed: One is that the first one or two tests of the series should be, if possible, of such a nature that they are not likely to be greatly affected by timidity or excitement on the part of the subject. Most persons, when approaching a test situation for the first time, do so with more or less trepidation. If given a test involving difficult thought or delicate coördinations at the beginning of a test series, such subjects may not give at all normal scores. The results of observation in this respect are confirmed by the fact that heart rate and blood pressure correlate negatively with difficult tests involving symbolic processes (page 154). Some investigators, such as Link, advocate the use of a dummy test of a rather simple nature, placed first. Such a test quite appropriately has been called a "shock absorber." This plan has not found general favor, probably because of the time consumed by it, though its principle seems sound. In such a case, of course, the subjects must regard the shock absorber as a genuine test or it will fail of its purpose.

A second principle of arrangement is that two tests should not be consecutive where the activity of the first would, through the principle of perseveration, be likely to persist into the activity of the second and thus cause an interference of one activity by the other.

EXPERIMENTAL CONDITIONS

As regards the general conditions under which the tests of the preliminary battery are to be given, it may be said that in so far as it is possible they should duplicate those under which the final battery will be administered. This applies both to the physical and the psychological environment and to the details of technique. For example, if the final battery is to be used near the roar of machinery, then the preliminary battery should be given under such conditions. If the final battery is to be used to test people singly, the preliminary battery should be administered to subjects singly, even though it might be susceptible of group application.

SECURING COÖPERATION FROM THE TRIAL SUBJECTS

When aptitude tests are administered to applicants for jobs or to persons seeking vocational guidance, there is practically never any difficulty arising from lack of coöperation on the part of the subject. The reasons are obvious. But the problem of securing the whole-hearted coöperation of the subjects of the trial group is not quite so simple. As a rule, however, little difficulty is encountered even in the latter case. If the subjects are employees, already on the job, they will naturally have nothing to gain by the experiment and they should be paid for their time at the regular rate. Subjects found in classes or otherwise in the charge of instructors usually give good coöperation through the favorable influence of the person in charge of the class. With the better-educated groups an additional appeal based on the need of the world for aptitude tests that may prevent others from making mistakes in choosing a life work, is very effective. Perhaps most effective of all is the almost universal desire on the part of subjects to make as good a score as possible.

Occasionally, however, the greatest difficulty may be encountered in securing the coöperation of subjects. The writer once attempted to administer some tests to the pupil nurses in a nurses' training school, run in connection with an insane hospital. The investigation had the full sanction and coöperation of the medical staff. Yet after one or two subjects had been through the tests it became impossible to get any others to enter the testing room except by compulsion. The most violent emotional reactions were manifested. It happened that the same tests had also been given during the preceding weeks to a large number of the insane patients. It developed upon investigation that some one had spread the idea among the nurses that they were being examined to discover whether some of *them* were insane! It is probably not without significance that a large number of male attendants in the same institution submitted to the same tests immediately after without a single protest or other indication of unwillingness.

As pointed out above in another connection, subjects often approach a test with a certain amount of trepidation. This is usually less where tests are given to groups, which fact is not the least of the advantages of group testing. In individual testing it is usually sufficient to engage the subject in conversation concerning some matter likely to interest him. This serves both to establish a friendly relation with him and at the same time largely to dissipate his fears of failure. In group testing a statement of the purposes of the test and of the desire of the experimenter that *all should do as well as possible* usually serves to relieve the tension sufficiently for practical purposes.

THE TECHNIQUE OF TIMING

An important detail in the conduct of most aptitude tests is that of timing. This may be either the task of giving

the starting and stopping signals for tests having a time limit, as in most group tests, or of taking the exact time consumed by a subject in performing a given task, as is frequently found in individual tests. In either form of timing a split-second stop watch (Fig. 43) should be used. These watches are so



FIG. 43. Showing a very useful yet reasonably priced stop watch.

constructed that they will start from zero at a pressure of the hand and will stop at a subsequent pressure, thus permitting the time elapsing between the two pressures to be read off. Thus the watch is pressed when a subject begins his test and again when he finishes. The watch then shows the time used by the subject in performing the test. A third pressure throws the hands back to zero in readiness for the next test. Such watches usually measure time to fifths of a second, which is quite close enough for practical purposes.

While for all kinds of testing work stop watches are much more convenient and their use is less likely to lead to error, it is possible, with care and with tests in which the time limits

are fixed, to use an ordinary watch having a seconds hand. This is because such time limits are usually placed at even minutes, or, at the worst, half-minutes. These can easily be noted on common watches. When an ordinary watch is used for such timing, the watch should be set in such a way that when the seconds hand falls at 60 the minute hand will fall exactly on one of the minute divisions of the dial, thus making the two hands agree. Then the experimenter should always be careful to give the starting signal exactly on the even minute or half-minute, which assists greatly in avoiding errors in timing. Lest an error should be made through the experimenter forgetting the exact position of the hands at the beginning of the interval, it is well for him to record on a convenient piece of paper, immediately after the starting signal, the exact time at which it was given and also the exact time at which the stop signal is to be given. For example, if a four-minute time-interval begins at nine forty-three, there would be recorded on the card :

Begin

9 : 43

Stop

9 : 47

With all kinds of watches, as the termination of the time interval approaches, the experimenter should hold himself alert, with his eyes on the hands, so as to give the stop signal at exactly the right instant. A few seconds' error in the timing of a test may play utter havoc in the interpretation of test results. The very greatest care should be taken that the timing is exact. The amateur is perhaps in greatest danger of error at this point of any in the giving of tests.

THE DURATION OF TESTS AND BATTERIES

In determining the time to be allowed for tests having a time limit, the principle to be followed varies according to the nature of the test. In the case of simple repetitive tests

where the nature of the activity is such that it may be continued indefinitely without running short of test material (e.g., tapping, steadiness, etc.), the time limit should be set at such a point that the reliability is as high as possible consistent with the time available. In the case of tests made up of a limited number of items such as Army Alpha, the time limit should be set at a point that will not quite permit the most rapid subjects to finish the material. If the time is much shorter than this, the last items of the series will never be used and will therefore be wasted. If it is much longer than this, a number of individuals will finish before the time is up and at unequal times. There will result a tendency to a lack of differentiation of the abilities of these fast individuals except as it may be revealed indirectly by the *quality* of their performance.

It is never safe to guess at the time limits to be allowed for tests with the trial group of subjects. If this is done, the chance is high that it will be found to have been either too long or too short. Such an error is a gross violation of the principle laid down earlier in the chapter; viz., that the conduct of the tests in the preliminary test battery should be in every possible way the same as when later administered in the final battery. If later an attempt should be made to include the test in the final battery, but with a corrected time limit, the whole series of ills will be encountered which were described above when discussing a similar change in directions or other experimental technique (page 344). This whole series of difficulties may be avoided very easily by the expedient there pointed out; i.e., of administering the tests to a pre-trial group of comparable subjects. By observing carefully the behavior of the more speedy subjects of this pre-trial group as the estimated time limit is approached, the correct time limit may be determined with precision.

Owing to the fact that the preliminary test battery is almost certain to contain a large number of test units which will not find a place in the final battery, the total time required of the trial group is nearly always excessive as compared with that for a final battery. For example, if it is desired to secure a final battery which can be administered in about 40 minutes, the time required for administering the preliminary battery might easily be 90 minutes or even more. This comparatively long time required of the trial group sometimes acts as a difficulty in the way of securing trial subjects. Unfortunately this is inherent in the method and cannot be avoided.

In case the whole or a part of an aptitude-test battery is of the pencil-and-paper variety, the various tests may profitably be stapled together into a convenient booklet. The front page should provide a blank on which the subject may write his name and other information which later may be of value, such as age, schooling, etc. In case paper tests are given separately instead of in a booklet, care must be taken that the subject has his name written on *each sheet* for purposes of identification in case the records accidentally get mixed. Where the performance of subjects on apparatus is to be recorded, special blank forms should be provided, with space for the names of the subjects and convenient places for their scores according to the nature of the test. Because of its stiffness it is often convenient to have these forms drawn up on a good quality of white cardboard, particularly when administering the tests to a trial group.

THE PROBLEM OF SCORING TEST RESULTS

After the tests have been administered to the trial group of subjects, there comes the important problem of scoring the test results. There is, of course, no difficulty in the case of tests having a simple time score, since this is read off

directly from the face of a stop watch at the completion of the test. The same is true of many apparatus tests which may be designed in such a way as to show the subject's score automatically by means of mechanical counters and other devices of a similar nature. On the other hand, many apparatus tests are so designed that there remains after the test some kind of a graphic record representing the subject's behavior, which must later be scored. An example of such a test is found in the Wisconsin miniature test for engine-lathe aptitude (Fig. 11). A description of the use of this apparatus has been given above (page 67). As the metal point is made to move around the series of contacts, an extension of the arm bearing this point makes on a square piece of paper an exact pencil tracing of the path actually followed. Now the direction to the subject is to pass from each contact straight to the next. A perfect score on this phase of the test would accordingly be a series of straight lines connecting the various points on the tracing, corresponding to the several contacts. But in case a subject had considerable difficulty in manipulating the apparatus there will be numerous deviations from a straight line, which will lengthen the course actually traversed. Thus the distance traversed becomes a score or indication of the test performance. Such a record may be scored for distance with sufficient accuracy by means of a cartometer or map measurer such as shown in Figure 44. This device has a small tracing wheel which may be made to follow any irregular line. To this wheel is attached, by means of gears, a hand which indicates the distance traversed on a suitable dial.

SCORING PENCIL-AND-PAPER TESTS BY MEANS OF KEYS

Since pencil-and-paper tests are nearly always group tests, they are usually given by the time-limit method and consequently must be scored. This, together with the fact that

they are so widely used, makes the matter of scoring them of considerable importance. We have already discussed



FIG. 44. Cartometer. (Reproduced by permission of Keuffel & Esser Company.) This instrument, designed for making map measurements, is very useful for measuring the length of irregular lines secured from test behavior.

(pages 309 ff.) certain problems relating to the speed and accuracy of the scoring of such tests, as related to the form of construction. We have now to consider how the process of scoring may be facilitated by the use of certain scoring devices.

Perhaps the most primitive scoring device for the pencil-and-paper test of the verbal variety is a key showing what the correct responses should be. This key, usually a sheet of paper containing the correct responses, is placed near the sheet being scored so that the eye of the scorer may glance back and forth from the one to the other as the scoring proceeds. This is naturally very tiring to the eye, to say nothing of the time consumed and the errors arising from inaccuracies in optical fixation. A number of special methods and devices have therefore been employed to minimize these difficulties. The simplest of these is for the scorer to memorize the key by rote at the outset of the scoring. Where a large number of test scores from a single test must be marked at about the same time this is frequently one of the very best methods.

In tests where the responses are brief and are arranged uniformly in a vertical column, the key may be made from a

narrow strip of cardboard upon which the true responses are placed at intervals corresponding exactly to the test form. This key can then be placed directly at the side of the column of responses on the test blank and any discrepancies between the two series can be detected instantly and with a minimum of eye strain. Frequently the same result is secured by having the key printed along one edge of a card. By utilizing all the margins of both sides of a single card, keys of this kind may be provided for as many as eight tests. The remainder of the card may contain brief directions for the use of the keys. An excellent example of this method of scoring is found in the Stenquist Mechanical Aptitude Tests.

VARIOUS TYPES OF SCORING STENCILS

In completion tests, particularly where the words to be written in are scattered in a more or less random fashion about the page, the above device is not applicable because the relevant parts of the key cannot be placed sufficiently close to many of the items which need to be scored. In such cases a very effective scoring device may be constructed from a light sheet of cardboard cut exactly the size of the sheet to be scored. In this cardboard there is cut a series of holes — one over each place that a word is to be written in by the subject. On the cardboard, just above each hole there is written distinctly for instant comparison, the word which should appear within the aperture. This scoring stencil has the advantage of concealing from the eye of the scorer the remainder of the page which, for him, is quite irrelevant and can serve only as a distraction.

A second device, similar in many respects to the one just described, is a transparent waxed paper upon which the key words are printed at appropriate places. Still a third device is a transparent sheet of celluloid upon which the key words

are etched or engraved at appropriate positions, as in the case of the waxed paper. Because of its stiffness the celluloid is much more convenient to handle than the waxed paper. The latter must be very thin, and also easily becomes wrinkled and torn by use. As a rule, transparent scoring stencils are to be preferred to opaque ones such as may be made of cardboard, only when the *position* of a mark made by a subject is significant and where errors are to be scored. A case in point is the multiple-choice test shown on page 311. Obviously, if a test of this kind were to be scored by an opaque stencil, it would be impossible to see any but the correct responses and the errors could not be scored.

The various advantages of both the opaque scoring stencil and the transparent type may be combined by simply perforating sheets of ordinary transparent celluloid as described above in the case of cardboard. Such a transparent stencil, when placed over a test sheet for scoring, reveals both successes and errors, but in a perfectly differentiated manner. The correct responses all appear clearly through the apertures, whereas the erroneous responses may be seen through the slightly yellowish medium of the celluloid. Such stencils may easily be made by placing the sheet of celluloid over a test form correctly marked, the two being held firmly together by a few drops of glue. The combination is then placed on a smooth block of hard wood. A hollow punch of the shape and size of the holes to be made is placed over every marked place on the form and given a sharp tap with a hammer. The celluloid easily fractures around the outline of the area bruised by the punch and may easily be pushed out. The resulting stencil should then be engraved with a sharp instrument to indicate which surface is uppermost in use and which edge goes at the top of the page. The sheet of celluloid should be the exact size and shape of the test form

to insure the apertures being placed rapidly and accurately in their true positions.

MACHINES FOR SCORING MULTIPLE-CHOICE TESTS
AUTOMATICALLY

Pressey has recently exhibited an ingenious machine into which multiple-choice material of the pencil-and-paper variety may be placed, the choices of the subjects being made by pressing keys. The machine counts automatically and separately the correct choices and the errors, both of which may be read off directly from dials at the conclusion of the test.

It is probable that the future will see group tests of the multiple-choice variety which will be scored entirely automatically by specially designed scoring machines. The designs for at least two such combinations are already in existence. Perhaps the most promising form of test response for mechanical scoring would be to have the subject make perforations in the test sheet given him, the position indicating his choice of the alternatives offered in the test. A machine could easily be constructed that would not only count up automatically both the right and the wrong responses in each test but which would also combine the various scores into an aptitude forecast, giving each its proper weight according to a scientific scoring formula. Such mechanical methods would not only reduce immensely the cost of scoring and computing aptitude forecasts but would also reduce the element of error resulting from human inaccuracy.

TRIAL SUBJECTS, FREE-HAND-DRAWING APTITUDE PROJECT

Having concluded our general account of the administration of the preliminary test battery to a trial group of subjects, we shall now describe in a little detail the actual procedure in the case of the free-hand-drawing project, the preceding steps of which we have already traced (pages 325 ff.).

The trial subjects were secured from a course in perspective drawing at the University of Wisconsin. The first step in securing the coöperation of these persons was to explain to the head of the department and the instructors in charge of the various sections the desirability of forecasting drawing aptitude. It happened that the department had already been considering the possibilities of a test for drawing aptitude, so that they at once fell in with the plan and gave the most whole-hearted coöperation throughout. They not only turned their classes over to the experimenters so that the preliminary battery could be administered to them, but gave invaluable assistance in the preliminary analysis of the drawing activity and especially in the securing of the criterion score. This latter was assured at once by the plan of having each subject, at the conclusion of the course, make a representative line drawing of the same fairly complex object, the drawings to be ranked independently by the various instructors (see page 413).

At the outset there seemed to be a fairly adequate supply of subjects, as the combined enrollment of the various sections of the course amounted to over 75 students. It was found, however, that on the days on which the tests were given there were many absences, so that several were lost in this way. Then at the end of the course, when the criterion test of drawing the object was given, quite a number of others were also lost. The combined losses from these sources left, at the end, only 48 subjects with perfect records suitable for use. This number was somewhat smaller than had been anticipated and was much smaller than was desirable. Indeed, it is probably about the minimum worthy of serious experimental consideration. This phase of the present investigation is, however, quite typical of the difficulties encountered in aptitude-testing projects.

As to the character of the trial group of subjects, it may be

said that they were good in one respect and unsatisfactory in another. They were satisfactory in the sense that they were the same class of pupils that the final battery was designed to be used upon. They were unsatisfactory in that they had received training in the activity before the tests were given, since the testing was done some weeks after the course began. Still worse, some of the subjects had received a considerable amount of drawing instruction in the lower schools, while others had received little or none.

EXPERIMENTAL CONDITIONS, DRAWING-APTITUDE PROJECT

Previous to the date set for the giving of the preliminary battery of tests, the test materials were all prepared and in good working order. This included the preparation of the various designs drawn on white cardboard for group presentation, the assembling of the various materials which made up the test booklets, the attaching together of the various sheets making up the booklets, the formulation of the various oral instructions for the respective tests, and so on. Not only this, but after all was prepared, a few individuals constituting a pre-trial group were put through the whole series as a kind of dress rehearsal to make sure that nothing had been overlooked and that the experimenter himself had his technique thoroughly mechanized. As is almost invariably the case, this rehearsal revealed a number of serious defects in the technique, particularly as to the adequacy of the verbal instructions. It would seem next to impossible to anticipate all emergencies in the design of a test without a pre-trial rehearsal. With the various inadequacies of the technique thus revealed, remedial measures were taken at once and the preliminary battery was ready to administer to the trial group.

The test battery was found to be sufficiently brief to be administered during a single class period of 50 minutes.

Accordingly, when the students assembled for their drawing instruction on a certain day, the instructor introduced the experimenter to the class and briefly explained the purpose of the investigation. The experimenter then stood before the class and, in a manner calculated to put the subjects at their ease, amplified the explanation somewhat. The students were then arranged in the room in such a way as to be able to see the test materials as favorably and as equally as possible. The test booklets were then passed out, with the following oral instructions:

“Write your name on the first sheet. Do not turn over the sheet until I tell you. These experiments are an attempt to find a set of tests which will detect natural aptitude for drawing. Some people are able to draw divinely, and we are trying to discover what it is in their make-up which makes this possible. We, therefore, wish every one to make just as good a score as possible.”

The various tests were then administered as described above (pages 325 ff.). At the conclusion of the test the booklets were gathered up and filed for scoring.

TYPICAL TEST RESULTS AND METHODS OF SCORING,
DRAWING-APTITUDE PROJECT

In order to give the reader a concrete notion of scoring problems confronting the aptitude psychologist, we have reproduced a complete set of responses of a typical subject (No. 1) and will take up each in turn and describe how the various difficulties were met.

Test for Memory of Design

The test performance on memory of design of our typical subject is reproduced in Figure 45. As pointed out above (page 326), this particular design was chosen rather than numerous others which suggested themselves, because it

MEMORY FOR DESIGN

Reproduce design on this page.



4



4


$$\begin{array}{r} 3 \\ \hline 11 \end{array}$$

FIG. 45. The performance and some scoring details of subject No. 1 on test of memory for design.

possessed especially favorable characteristics for scoring. The method finally decided upon was to count each straight line found in the design as a unit of the score if reproduced and in approximately the right proportion. By comparing the reproduction of the design (Fig. 45) with the original (Fig. 38, page 328), it will be seen that the upper figure has all four strokes correct. The second figure also has all four strokes correct. The third figure is entirely lacking. The bottom group is present, but only the three vertical strokes are correct. The subject accordingly receives a score of 11.

Directions Test

The performance of subject No. 1 on the directions test is shown in Figure 46. The scoring of this test requires little comment. At places such as the dotted space in the fifth line, where the directions require *nothing* to be done, it was counted as correct in case no mark had been made there, provided a subsequent direction had been performed. The subject put a dot over the *H*, a comma after *mother*, left the space blank after the word *here* (as he should), put a *4* after the word *has* and a *yes* after the word *not* — all of which were correct. He therefore receives a score of 6 right and no errors.

Downey's Slow-Movement Test

The performance of subject No. 1 on the slow-movement test is given in Figure 47. An inspection of this scroll reveals that each cycle may be divided into an upward stroke and a downward stroke. In case a subject stops in the midst of a stroke, it was disregarded if it was less than a half and counted as a full stroke if more than half the stroke had been traversed. In the case of the record which has been reproduced, the pencil may be seen to have traced nearly six of the strokes. The score is accordingly taken as six.

With your pencil make a dot over any one of these letters F G H I J, and a comma after the longest of these three words: boy mother, girl Then, if Christmas comes in March, make a cross right here..... but if not, pass along to the next question, and tell where the sun rises...*in East*..... If you believe that Edison discovered America, cross out what you just wrote, but if it was some one else, put in a number to complete this sentence: "A horse has...*4*...feet." Write *yes*, no matter whether China is in Africa or not ...*yes*...; and then give a wrong answer to this question: "How many days are there in the week?"..... Write any letter except *g* just after this comma, and then write *no* if 2 times 5 are 10..... Now, if Tuesday comes after Monday, make two crosses here.....; but if not, make a circle here.....or else a square here..... Be sure to make three crosses between these two names of boys: George.....Henry. Notice these two numbers: 3, 5. If iron is heavier than water, write the larger number here....., but if iron is lighter write the smaller number here..... Show by a cross when the nights are longer: in summer?..... in winter?..... Give the correct answer to this question: "Does water run uphill?"..... and repeat your answer here..... Do nothing here ($5 + 7 =$ ), unless you skipped the preceding question; but write the first letter of your first name and the last letter of your last name at the end of this line:

L R

FIG. 46. The performance of subject No. 1 on the Woodworth-Wells directions test.

DOWNNEY GROUP WILL-TEMPERAMENT TEST

VII Trace as slowly as possible the scroll below.

1



2



FIG. 47. The performance of subject No. 1 on the Downey slow-movement test. (Published by World Book Company.)

Owing to the fact that this test might have been scored very much more precisely than as described above, it may appear at first sight that our method is unduly coarse. It must be observed, however, that the precision desirable in scoring is in part a function of the range of the behavior being scored. In this case it ranges between 2 and 54 points (X_4 , Table 77). If the attempt had been made to estimate the fractional parts of strokes traversed, there would have resulted at least an extra figure to be handled in the computations, which would have increased the labor appreciably. The added precision secured would not have been worth the trouble.

Reproduction-of-Angles Test

The reproductions of the angles shown in Figure 39 as made by subject No. 1 are given in Figure 48. The point of this test was to show how closely a person could reproduce the angles. Accordingly it was necessary, in scoring, first to measure the angles drawn by the subject. This was done by means of a small transparent celluloid protractor. By means of this instrument angles could be measured to a half degree with considerable accuracy. The difference in degrees was then found between the original and the reproduced angle. These five angular differences were then added, regardless of whether they represented overestimates or underestimates on the part of the subject. The total angular error was the final score. In order to avoid handling large decimals in the final score, all fractions of $\frac{1}{2}^{\circ}$ or over were counted as a full degree and all of less were dropped. In the case of the present subject, his first angle measured 32° , whereas the original angle was 37° , which yields an error of 5° . The second angle drawn measured 46° as compared with 49° for the original, thus showing an error of 3° ; and so on with the rest.

A certain amount of difficulty was experienced in measuring the reproduced angles owing to the fact that the lines bounding them were often not quite straight. In such cases the angle was read as nearly as possible to what it would have been if bounded by the best-fitting straight line. Further difficulty was caused by the fact that in some cases the lines for a given angle would come together sharply and that in others they would end bluntly, thus making it difficult to center the protractor. In the latter cases the true point was estimated as accurately as possible before placing the protractor.

REPRODUCTION OF ANGLES

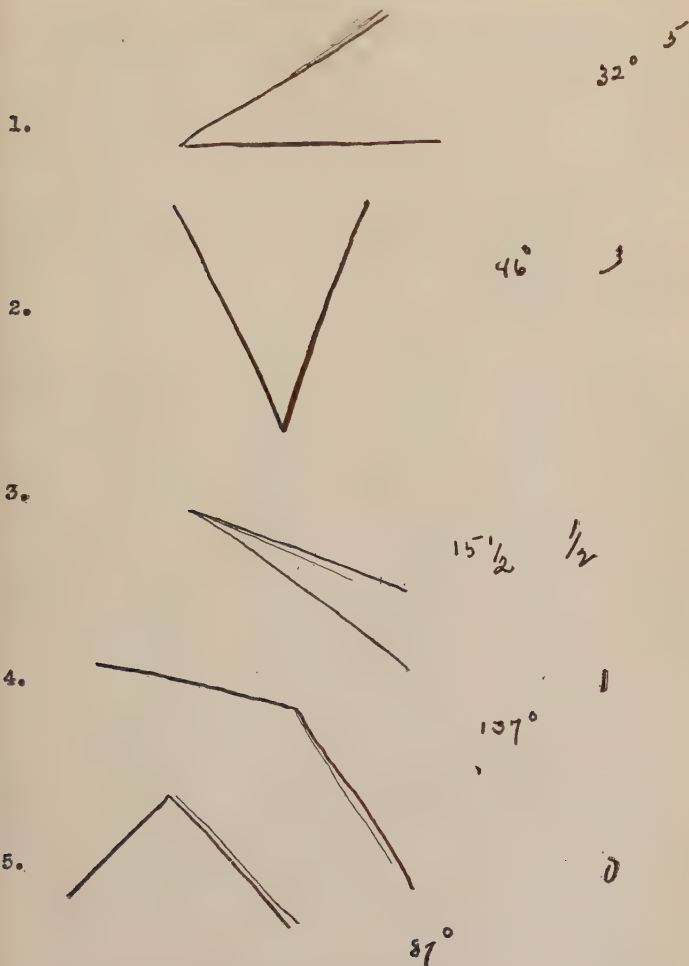


FIG. 48. Showing the performance of subject No. 1 on the test for the reproduction of angles.

Opposites Test

The performance of subject No. 1 on the opposites test is reproduced in Figure 49. It may be seen that he skipped one item and marked two others incorrectly. Eight items were performed correctly; so his score is 8. A simple key for scoring this test may be constructed by filling out a blank form of the test correctly and then with scissors severing the two response columns from the rest of the form. This may be placed alongside of a record to be scored and greatly facilitate the process.

Judgment-of-Size Test

The responses of subject No. 1 in the judgment-of-size test are shown (greatly reduced) in Figure 50. The problem of scoring these responses presented several alternatives. Since it was avowedly a test of ability to reproduce size, it was decided not to deduct anything for wrong proportions as such. There is also the possibility of considering the absolute area. The most accurate method of measuring areas of this kind is by means of the planimeter (Fig. 52). Unfortunately this method is very time-consuming. It was finally decided to measure certain of the transverse dimensions of each figure and take the sum of the errors as the score. Thus the circle was $3\frac{6}{16}$ inches in diameter. The "circle" drawn by our subject was $3\frac{7}{16}$ inches in vertical diameter and $3\frac{2}{16}$ inches in horizontal diameter. This yields an excess of $\frac{1}{16}$ inch on one diameter and of $\frac{6}{16}$ inch on the other, or a total error of $\frac{7}{16}$ inch. Corresponding measurements were taken on each of the other figures, the total of all the errors on all the figures being taken as the final score. In the case of subject No. 1, these totaled $\frac{25}{16}$ or approximately 1.6 inches, which was taken as the final score.

About the only difficulty encountered in making these

Opposites

(Prepared by Prof. Carl C. Brigham, Princeton University)

Each group of four words in the thirty lines below contains two words which are either (a) the same or nearly the same in meaning, or (b) the opposite or nearly the opposite in meaning.

Find the two words in each group that are either same or opposite, and write the numbers of these two words in the column at the right, headed "Same," or the column headed "Opposite," as the case may be.

The first group of words, "1 bent, 2 cold, 3 hot, 4 sad" contains two words ("cold" and "hot") that are opposite in meaning, so that the figures 2 and 3 are entered in the column headed "Opposite." The second and third groups have also been marked correctly.

				SAME	OPPOSITE
1 bent	2 cold	3 hot	4 sad	<u>8</u>	<u>2 & 3</u>
1 white	2 safe	3 black	4 raw	<u>8</u>	<u>1 & 3</u>
1 rapid	2 hard	3 large	4 great	<u>3 & 4</u>	<u>8</u>
1 wet	2 dry	3 thin	4 easy	<u>8</u>	<u>1 & 2</u>
1 flat	2 tired	3 smart	4 level	<u>1 & 4</u>	<u>8</u>
1 corpulent	2 opulent	3 torpid	4 affluent	<u>8</u>	<u>8</u>
1 lavish	2 jaunty	3 gradual	4 demure	<u>8</u>	<u>2 & 4</u>
1 partial	2 careful	3 jubilant	4 exultant	<u>3 & 4</u>	<u>8</u>
1 animated	2 vivacious	3 tarnished	4 caustic	<u>1 & 2</u>	<u>8</u>
1 formidable	2 tangible	3 preferable	4 alarming	<u>8</u>	<u>1 & 2</u> ^L
1 captivating	2 belligerent	3 pacific	4 universal	<u>8</u>	<u>3 & 4</u> ^L
1 capacious	2 petrified	3 obligatory	4 restricted	<u>8</u>	<u>1 & 4</u>
1 agile	2 lithe	3 hirsute	4 previous	<u>1 & 2</u>	<u>3 & 4</u>
1 bombastic	2 obnoxious	3 astute	4 modest	<u>8</u>	<u>1 & 4</u>
1 maudlin	2 obvious	3 mawkish	4 oblique	<u>8</u>	<u>8</u>
1 lingual	2 raucous	3 overt	4 manifest	<u>8</u>	<u>8</u>
1 delectable	2 mutable	3 reflexive	4 permanent	<u>8</u>	<u>8</u>
1 literate	2 choleric	3 irascible	4 ambidextrous	<u>8</u>	<u>8</u>
1 illegible	2 agnostic	3 amenable	4 refractory	<u>8</u>	<u>8</u>
1 punctilious	2 brusque	3 translucent	4 insensitive	<u>8</u>	<u>8</u>
1 legitimate	2 disparate	3 unequal	4 metric	<u>8</u>	<u>8</u>
1 enigmatic	2 extraneous	3 loquacious	4 intrinsic	<u>8</u>	<u>8</u>
1 laconic	2 dulcet	3 epigrammatic	4 titanic	<u>8</u>	<u>8</u>
1 gratuitous	2 sanctimonious	3 rhythmic	4 obligatory	<u>8</u>	<u>8</u>
1 anomalous	2 efficacious	3 jugular	4 usual	<u>8</u>	<u>8</u>
1 jovial	2 nascent	3 incipient	4 nauseous	<u>8</u>	<u>8</u>
1 capricious	2 spectral	3 speculative	4 corporal	<u>8</u>	<u>8</u>
1 pertinent	2 prismatic	3 congruous	4 juridical	<u>8</u>	<u>8</u>
1 infantile	2 infinitesimal	3 amorphous	4 shapeless	<u>8</u>	<u>8</u>
1 nugatory	2 efficacious	3 gregarious	4 transcendental	<u>8</u>	<u>8</u>

Stop here. Wait for further instructions.

FIG. 49. The record of subject No. 1 on the Brigham opposites test.

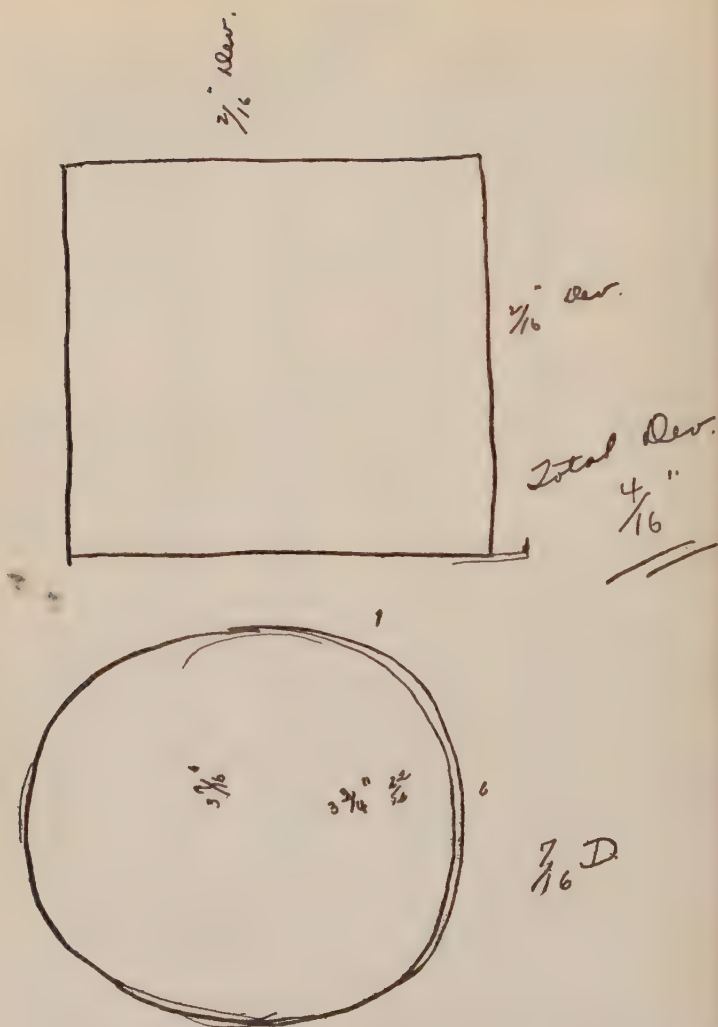


FIG. 50 a. Showing performance and scoring notations of subject No. 1 on test for judgment of size.

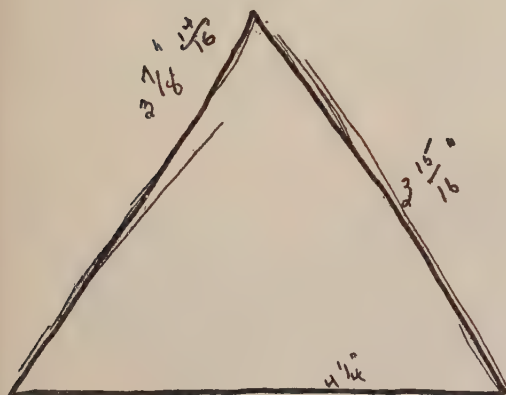
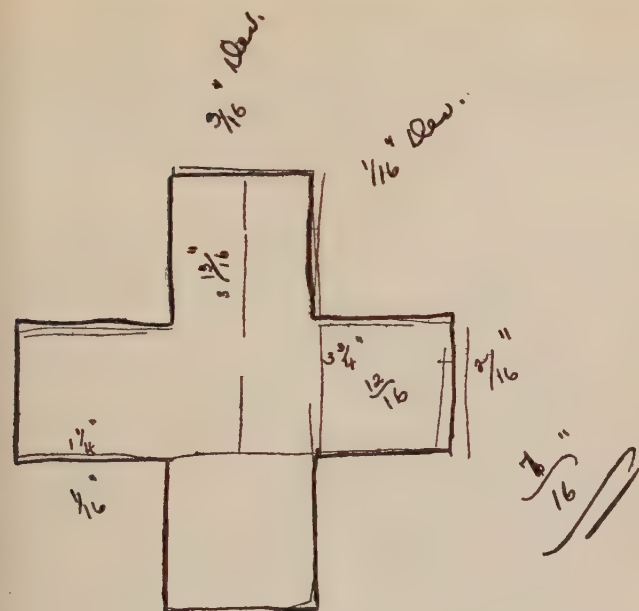


FIG. 50 b. See Figure 50 a.

" T.D.
 $7\frac{1}{16}$
 3.

CIRCLE-COMPLETION TEST

In this experiment you will be allowed $\frac{1}{2}$ minute to complete each circle. When the signal is given, trace with your pencil the rest of the circle, making it complete. Do not start on circle 2 until you are told. Make the circles as perfect as possible.

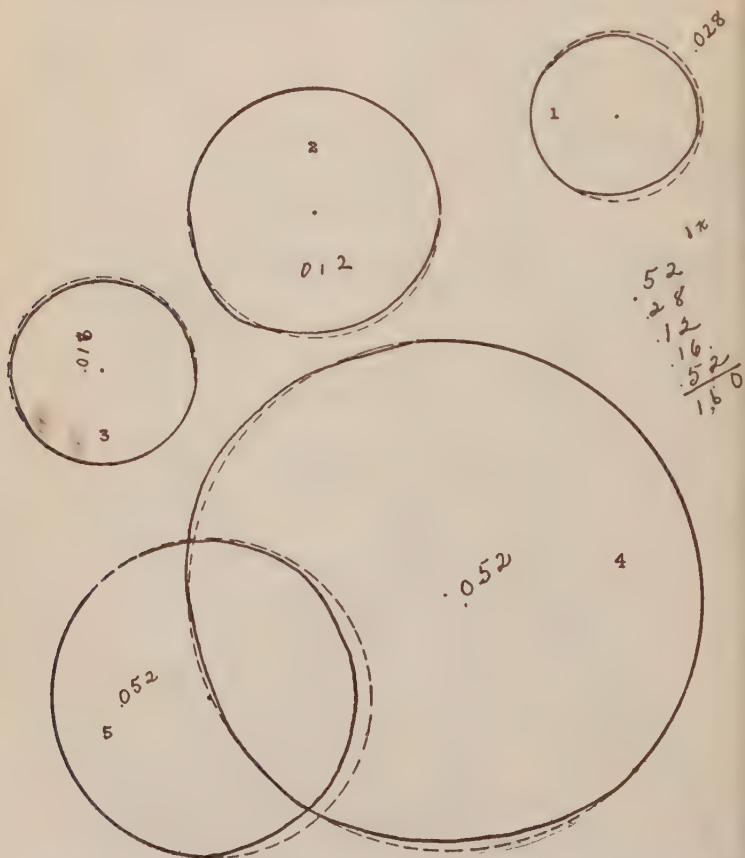


FIG. 51. Showing the record of subject No. 1 on the circle-completion test. The dotted line represents the true circle completion.

measurements was in cases like that shown in the circle figure of Figure 50 *a*, when the subject made two or more lines where there should be but one. In such cases the scorer took the line which looked as if it were the final action of the subject.

Circle-Completion Test

The test responses of subject No. 1 in the circle-completion test are shown in Figure 51. The blank form of this has already been shown above (Fig. 41). The scoring of this test presented the greatest difficulty of any in the series. Fortunately, in the mimeographing of the test form a minute dot was left to mark the true center of each circle. By means of this a fine compass could be set so as to draw the true completion to the circle (shown in the figure by a broken line). The discrepancy between the line made by the subject and the true circle then became manifest. The question then arose how the extent of this discrepancy could be measured. This was solved very readily by the use of the planimeter, a special instrument for measuring areas of irregular contour (Fig. 52). The needle at the end of one of the legs of the instrument is pricked down firmly into the paper, and the needle at the other end is made to trace the lines bounding the

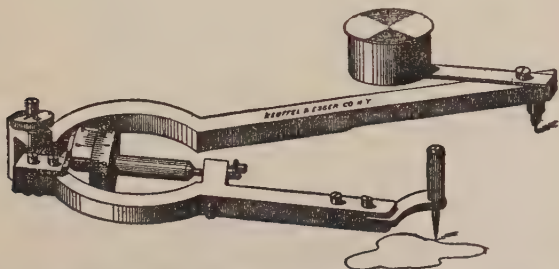


FIG. 52. Polar planimeter. (By permission of Keuffel & Esser Company.)
This instrument is invaluable for measuring areas of irregular contour.

First take your pencil and make a dot on the vertical line A' where horizontal line A would intersect if it were to be extended. Then do the same with the lines of B , C , D , and E . Remember, do not draw any lines or use any straightedge or sight along the lines. Locate the dots as accurately as possible.

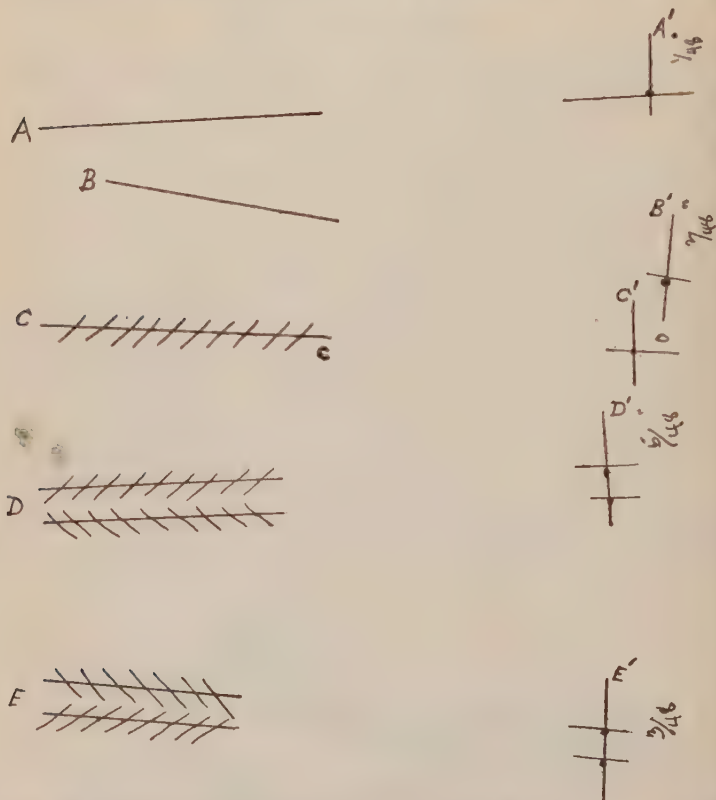


FIG. 53. Showing the record and scoring notations of subject No. 1 on the test for line extension.

discrepancies between the true and the attempted reproductions. When the tracing is complete, the area of the enclosed space in decimals of a square inch may be read off from a small wheel supporting the legs of the instrument near their point of junction. The discrepant areas of each circle completion were measured, and all five were added together for the total score. Because of the small total of discrepancy the decimal of the final score was set one place to the right, so that the final unit of measurement was $\frac{1}{10}$ of a square inch.

The Line-Extension Test

The test responses of subject No. 1 on the line-extension test are shown in Figure 53. The blank form of this test has been shown above (Fig. 42). The scoring of this test proved quite simple. A straightedge was carefully placed on each line to be extended, and a short straight line was then drawn through the vertical lines upon which the subject had made his dots. If the subject made no error in the placing of his dot for a given line extension, the line drawn by means of the straightedge should pass through the center of the dot. This is practically the case with the dot at *C*. The amount of deviation or error in the placing of the dot was then measured by placing over the region to be measured a transparent scale which could be read to $\frac{1}{48}$ inch. This fraction of an inch was accordingly taken as the scoring unit. The first dot shows an error of $\frac{1}{48}$, the second of $\frac{2}{48}$, the third of 0, and so on. The total amount of error on all seven lines was $\frac{1}{48}$ or .23 inch, which is the final score.¹

¹ The test scores of all of the subjects are assembled in Table 77, page 440. There and throughout the remaining chapters, X_1 represents Memory for Design; X_2 represents Directions, Rights; X_3 represents Directions, Wrongs; X_4 represents Downey Slow Movement; X_5 represents Reproduction of Angles; X_6 represents Opposites; X_7 represents Judgment of Size; X_8 represents Circle Completion; and X_9 represents Line Extension.

CHAPTER TWELVE

THE DETERMINATION OF THE ACTUAL APTITUDES OF THE TRIAL SUBJECTS

As a rule the most formidable problem encountered by the aptitude psychologist is the location of a trial group of subjects from whom a valid and reliable quantitative criterion of aptitude may be obtained. The difficulties of securing such criteria, coupled with the fact that to proceed on a scientific aptitude project without an adequate criterion is hopeless, have stimulated psychologists to devise various techniques by which these obstacles may be met. The present chapter will accordingly be devoted to the description of the numerous experimental and statistical procedures and methods which may be utilized in this fourth and most difficult step of an aptitude-testing project.

A satisfactory criterion score must be a numerical expression of the position that a more or less specialized aspect of a subject's behavior takes on a linear scale. It is to be observed that on a true linear scale a step or unit (as an inch, a pound, a degree of arc) taken from one part of the scale must be equal to a step or unit taken from any other part of the scale. Most psychological tests satisfy these conditions reasonably well. This is because they have been designed especially for measuring purposes. Occasionally, a criterion scale also may be obtained quite simply and directly. But owing to the great variability of human behavior, unless carefully controlled the conditions for accurate measurement of occupational activities often can be brought about only in a somewhat indirect, complicated, and imperfect manner. Unfortunately, in contrast to test behavior, the activities which constitute vocations have been determined by their practical utility rather than for the metrical convenience of aptitude psychologists.

THREE TYPES OF APTITUDE CRITERIA

The criteria upon which the ultimate aptitudes of the individuals of trial groups of subjects are based fall naturally into three classes. By far the most desirable criterion is some objective product resulting from the occupational activity of a subject working through a given length of time. Examples of such products are the bricks laid by a mason, the buttonholes made by a seamstress, the words transcribed by a typist, the paper boxes made by a factory worker, etc. This will be called the *product criterion*.

It happens that some occupational activities either do not leave any permanent objective results or, if they do, the results are not of such a nature as to be accurately measurable. This gives rise to a second class of aptitude criteria. In this, the scoring of the occupational behavior is accomplished by the direct observation of the occupational activity itself as distinguished from the product of activity. Thus the performance of an athlete in running the mile ordinarily does not have any measurable product. His activity must therefore be observed directly and timed with a stop watch as it takes place. Similarly, the rendering of a song must actually be heard to be scored for merit, and so on. This second type will be called the *action criterion*.

But frequently, owing to the irregularities in the conditions under which the individuals of the trial group pursue their activities, neither their behavior nor the products of their behavior are directly comparable, or even measurable, by ordinary methods. Or it may be that the data were originally satisfactory enough but were not preserved. This brings us to a third type of aptitude criterion. Often, even under such unpromising conditions as described, it is possible to secure very useful criterion scores by having the subjects *rated* by one or more persons who have seen a great deal of

their work. The persons who do the rating are usually foremen or other supervisors with an intimate knowledge of the circumstances under which the individual subjects work, who are therefore able more or less accurately to allow for special handicaps or advantages. It is to be observed that this third class of criteria resembles the first in being based upon a result or product of the occupational activity of the person rated, but that the result differs in being an impression made on the organism of the rater rather than on some object like a brick or a piece of cloth. In this class of criteria the rater in reality is rating the aggregate impression left in his own organism by the subject's various occupational activities which have been observed by him. Such criteria and ratings are thus essentially subjective rather than objective. This third class of criteria will accordingly be called the *subjective impression criterion*.

Because of the frequency with which such subjective criteria need to be employed in securing criterion scores, a great deal of ingenuity has been directed to perfecting the technique. It is worthy of note that a very respectable degree of success has been attained. This matter will be taken up again later in the chapter.

FIVE METHODS OF SCORING CRITERIA

In conjunction with the three types of criteria there must be considered five fairly distinct methods by which the scoring of various criteria may be accomplished:

The first method of scoring criteria is that of *simple counting*.

The second is that of measurement in terms of the *objective scales* and units of physical science such as feet, pounds, and seconds.

The third is by means of *subjective scales* such as those used in assigning school marks by the traditional scholastic-marking system.

A fourth is by means of serial arrangement in an order of merit. This will be called the *ranking method*.

A fifth is measurement by means of *qualitative scales*, usually made up of samples of the thing to be scored. Handwriting is frequently measured for general merit by means of such a scale.

Not all of the five scoring methods find application on all of the three classes of criteria. The most conspicuous example of this exception is the third class of criteria (subjective impressions of the scorer), which obviously cannot be scored by either of the first two scoring methods — i.e., simple counting and measurement by means of objective scales.

SCORING CRITERIA BY SIMPLE COUNTING

As already suggested, the method of simple counting has a rather limited usefulness in scoring aptitude criteria. For it to be of value with the action criterion, the activity in question must be made up of recurring homogeneous cycles which thus constitute natural units and accordingly may be counted. An example would be the motion cycles of a mason in laying bricks. The laying of each brick constitutes a cycle. These motion cycles may be counted, and the number of cycles performed in a given period might accordingly be used as an aptitude criterion. As a general thing, however, when the vocational activity runs in recurring cycles, the product of the activity also tends to fall into corresponding natural units which may themselves be counted. In such cases the product criterion is usually to be preferred to the action criterion. This is especially true where, as should usually be the case, a large sample is taken. It consumes as much of the scorer's time to score an action criterion as it does of the subject's time to perform the action itself. On the other hand, the number of bricks laid or the number of buttonholes made in a day can be counted in a few minutes.

One of the most common forms of criteria scored by counting is an educational achievement test, the units counted being the items of the test successfully performed.

SCORING CRITERIA BY OBJECTIVE MEASUREMENT

The method of scoring criteria by objective measurements is limited for the most part to product criteria. All things considered, this combination of criterion type and scoring method, where applicable, is the most satisfactory of any. The investigator should always employ objective measurement of product criteria if the conditions of the project permit. A description of the use of this method in a variety of practical situations may serve the double purpose of introducing the reader to some of the concrete details of securing criteria and of illustrating the use of this particular combination of criterion type and scoring method.

Dr. Henry C. Link undertook the task of organizing a battery of tests which would select efficient shell inspectors for an arms-manufacturing company. After a careful psychological analysis of the job, he located 52 girls all of whom were engaged in this occupation under practically the same conditions. After selecting his tests and administering them to this trial group, he sought a criterion score. This was found in the *average number of pounds of shells inspected per hour of work for the preceding month*. In such plants the employees are generally paid according to the amount of work performed, and consequently records of the amount done and of the number of hours spent are usually available. It appears that in this case the majority of the girls had been working the greater part of the four-week period on a single kind of shell. This was most fortunate. Link therefore threw out the records of all work on other kinds of shell so that the final average poundage should be strictly comparable

for the various workers. In many ways the above example represents an ideal criterion score.

It must not be supposed, however, that having a product criterion susceptible of measurement by ordinary physical methods insures a satisfactory criterion score. The experience of Howard Pollock is illuminating on this point. Pollock (68) undertook to devise a battery of tests to forecast aptitude for operating the knitting machines of a large hosiery-manufacturing company. Upon preliminary inquiry it was found that the company had in operation an elaborate system of records from which could be found the exact number of hours worked by each girl on any given day, the number of pounds of hose knit on that day, and so on. After analyzing the job, devising an elaborate preliminary battery of tests, and giving these tests to a trial group of 60 experienced knitters, he went to the company records to get his criterion score. This was to be the average hourly poundage produced during the preceding ten-weeks period. He found to his dismay that the 60 girls were working on several different types of knitting machines and were knitting no less than seven different styles of hose. On looking into the matter further, he discovered that some of the styles were men's hose, short and made from comparatively coarse cotton thread, and that others were women's fancy hose made from four or five silk threads knit simultaneously. In some cases these threads were so fine that it was necessary to pass them through an oiling device to prevent breaking during the knitting operation. There were too few girls working on any one kind of hose to make an evaluation of the tests possible on the basis of a single style group. Indeed, the entire group of 60 subjects was scarcely adequate even if all had been doing comparable work. It was therefore decided to reduce the production scores of the different style groups to the equivalent of a single style of women's hose. Such

reductions can be made with considerable precision (page 396). Unfortunately, any reduction to equivalent scores must assume that the groups to be made equivalent and comparable are, on the average, of approximately the same aptitude and the same range of variability. Upon investigation this was found not to be the case with the various knitting groups. The less apt individuals, as well as the less experienced ones, were placed on machines knitting men's hose, whereas only the more clever ones worked on the women's fine hose. To make matters still worse, the work of the company required the girls frequently to shift from one style to another, and it took a girl some time after shifting to reach her maximum production efficiency on the new style. The final result was that no adequate criterion score was secured and the project was a practical failure. Unfortunately, the conditions just described are not uncommon in industry.

The methods followed and the nature of the difficulties encountered in securing a satisfactory criterion score where there is a product criterion measurable by physical methods may be illustrated still further by an investigation carried out in the Engineering College at the University of Wisconsin. The problem was to organize a battery of tests that would be useful in forecasting ability to learn to operate the engine lathe as given in one of the important shop courses. During the course each man must turn out on the lathe a number of work-pieces. The production of these work-pieces constitutes the main work of the course. The men are furnished with blueprints which specify the exact dimensions of the work-pieces in detail. It is obviously a simple matter to go over these work-pieces with a micrometer and determine the amounts of deviations from the specifications. The smaller the amount of error or deviation, the better the score. In all, twenty-seven different dimensions were measured on the different pieces made by each subject.

The errors on each dimension were converted into roughly equivalent school grades. The method was as follows: There was first made for any given dimension measured, a distribution showing the number of subjects making each degree of error. By counting in from each extreme of this distribution it was possible easily to locate a point cutting off from the main distribution the 5 per cent of subjects making the greatest errors and the 5 per cent making the smallest errors. The error falling at the first point was given arbitrarily a grade of 70, and the error at the latter point was given a grade of 93. The various other errors in that dimension received grades in proportion to the scale thus set. The 27 dimensional errors, when thus converted into roughly equivalent school marks, were then averaged by a system of weighting based on the composite judgment of the instructors in the course as to the relative importance of the various dimensions as indicative of good workmanship. This average was the criterion against which the test battery was checked and upon which the weights of the tests finally chosen were based.

The conditions under which the above criterion scores for aptitude in learning the engine lathe were obtained were probably considerably better than are encountered on the average, yet they were by no means perfect. In the first place, the machines used by the men were not exactly alike. Secondly, it was known that occasionally men attempted to turn in work-pieces as their own which were really made by a student in the course a preceding year. A few such cases were detected by those in charge of the course and the guilty ones punished. Thirdly, and by far the most serious of all, was the fact that if a man turned in a piece which was too bad, he was given a new blank and required to do it over. In fact, many men would do certain pieces over several times before they were satisfied even to hand them in, working

overtime on Saturdays and holidays to do so. Thus a man with small aptitude but much industry might turn in a better product than a man of superior aptitude but moderate industry. The permitting of the men to use such grossly varying amounts of time to do their work, while doubtless good pedagogy, seriously injured what might otherwise have been nearly an ideal criterion score. This project serves, however, to illustrate once more the fact that occupational activities are determined by practical needs and not for the convenience of the psychologist.

SCORING CRITERIA BY RANKING

Leaving the field of objective measurement, we shall pass at once to the opposite extreme where there exists no objective scale or ordinary unit of measure whatsoever. To a person unfamiliar with the developments of modern psychological technique the securing of a valid quantitative criterion score under such conditions seems preposterous. The secret of the success of modern psychologists in this difficult task has been due largely to the discovery that in cases where no objective scale exists but where a number of different samples are to be scored at the same time, the scorer can usually arrange them in the order of the degree to which they possess a particular trait or kind of merit. This process of serial or order-of-merit arrangement is usually called *ranking*. The numbers which represent the ordinal positions of the various items so arranged are called "ranks." *In reality, what happens in the ranking technique is that the items to be scored are utilized as a kind of scale by which they themselves are measured.*

An illustration from a non-aptitude field may make this technique clear. Suppose it were desired to score the photographs of 50 young women for beauty. There is no objective unit for female pulchritude, to say nothing of a scale. Yet

most persons recognize very definitely different degrees of this trait — are able to state very definitely that one person is more beautiful than another. This power of discrimination between actual specimens, of telling which possesses more and which less of a given trait, makes it possible to arrange or rank them in a series on this basis. Accordingly it is possible after sorting the 50 photographs over for a few minutes, to pick out the one that seems to the rater the most beautiful, the one next most beautiful, and so on down to the one that is the least beautiful of all. Then the photographs may be assigned numbers, the most beautiful being No. 1, the next most beautiful No. 2, and so on down to the least beautiful, which will receive the number 50. *These numbers are the ranks.* It is usual, as in the case of the photographs just mentioned, to begin counting or numbering at that extreme of the series possessing the highest degree of the trait being ranked.

RANKS NOT TREATED LIKE ORDINARY UNITS

But it must never be forgotten that a series of ranks does *not* constitute a scale of ordinary linear units such as make up the scales familiar to physical science. It is not possible, for example, to equate the interval between any two successive ranks with that between every other pair of adjacent ranks. This limitation is inherent in the normal or Gaussian law of distribution. This type of distribution has been found to hold approximately in the case of nearly all biological traits which are measurable by objective physical methods. The probability that traits such as we now have under consideration and which happen *not* to be so measurable, also follow substantially the same general law, amounts almost to a certainty.

The general nature of the “normal” distribution may be seen at a glance in Figure 54. This represents by a series of

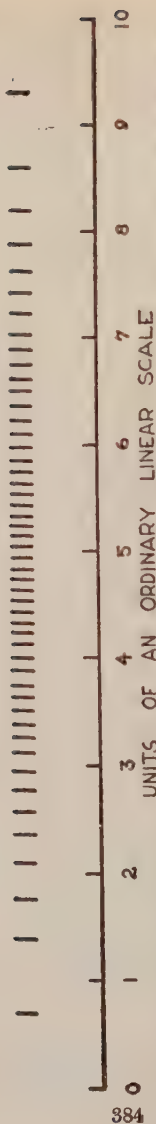


FIG. 54. Showing the variability in the absolute interval between successive ranks in a normal series of 50 individuals which are represented by the 50 vertical marks. Intervals between successive individuals in the middle are much smaller than at the two extremes.

50 short vertical marks the theoretically normal distribution of 50 individuals on a ten-point scale. It will be observed that the distance between the first and second marks at the extreme right (ranks 1 and 2) is many times as great as that between the twenty-fifth and the twenty-sixth marks (ranks 25 and 26) at the middle of the distribution. It is thus perfectly evident that ranks must not be treated like ordinary units. It is consequently necessary to exercise great care in using them. They should never be averaged, for example, except where the very roughest approximations are desired.

For substantially the same reason the Pearson product-moment correlation coefficient (r) (page 423), which must be used as a basis of the final evaluation and weighting of the test battery, cannot be computed directly from rank scores. It is true that several different methods have been devised by which correlation coefficients of a sort may be secured from ranks. The most notable of these is the Spearman rank-difference method. This method attained at one time considerable vogue. Even perfectly good linear scores were sometimes converted into ranks, in order to use this clumsy and inefficient method of calculation. Fortunately its use seems to be distinctly on the wane. Even where one or both variables to be correlated are rank scores, methods are now available whereby Pearson coefficients may be obtained without difficulty.¹

Quite apart from the lack of equality in the intervals of ranked scores is the fact that the numerical size of the largest scores of a ranked series has little or no significance except as indicating the size of the group which was being ranked. Thus the least comely girl in a small group might have a rank score of 10, whereas the same girl in a large group might receive a rank score of 100. This obviously makes it impossible to combine directly the gross ranks of several of the

¹ See pages 386 ff.

groups of different size in order to make up a single large experimental group. It is often necessary in aptitude work thus to combine groups in order to secure an adequate trial group of subjects. For the same general reason aptitude forecasts made in terms of ranks also would mean little or nothing. A predicted score of 50 might mean quite the worst possible score if the trial group chanced to have been composed of only 50 individuals, whereas it might mean just average or even much better than average if the trial group chanced to have been composed of 100 or 200 individuals. It is thus quite clear that ranks cannot be used in their original form as final criterion scores.

CONVERSION OF RANKS INTO ORDINARY SCALE UNITS

Fortunately the numerous and serious limitations on the use of rank scores noted above may be avoided in a very simple manner. This is accomplished merely by transmuting the ranks into their equivalent linear scores on some conventional scale. A convenient scale for this purpose is one of ten points. The nature of the change brought about by transmuting ranks to a 10-point scale of ordinary units may be seen very readily by glancing once more at Figure 54. There it may be observed that the individual of first rank receives a score of 9.3, that the person ranked 2 receives a scale score of 8.6, and so on. The transmutation is brought about by means of a simple formula and the use of Table 63. By these, any series of ranks can be translated readily into corresponding linear scores. The method assumes (37) that the trait in question has a "normal" distribution. We know, of course, that in the case of any limited number of data such as are found in ranked series, this assumption is never exactly true; yet there is reason for believing that practically always with good samples of human behavior it is approximately true. In practice the method yields excellent results.

TABLE 63

FOR TRANSMUTING "PER CENT POSITION" (SEE TEXT) IN RANKED SERIES INTO SCORES OR UNITS OF AMOUNT ON AN ORDINARY SCALE OF TEN POINTS. IF THE DECIMAL POINT IS DISREGARDED, THE SCALE BECOMES ONE OF 100 POINTS.

PER CENT POSITION	SCALE SCORE	PER CENT POSITION	SCALE SCORE	PER CENT POSITION	SCALE SCORE
.09	9.9	22.32	6.5	83.31	3.1
.20	9.8	23.88	6.4	84.56	3.0
.32	9.7	25.48	6.3	85.75	2.9
.45	9.6	27.15	6.2	86.89	2.8
.61	9.5	28.86	6.1	87.96	2.7
.78	9.4	30.61	6.0	88.97	2.6
.97	9.3	32.42	5.9	89.94	2.5
1.18	9.2	34.25	5.8	90.83	2.4
1.42	9.1	36.15	5.7	91.67	2.3
1.68	9.0	38.06	5.6	92.45	2.2
1.96	8.9	40.01	5.5	93.19	2.1
2.28	8.8	41.97	5.4	93.86	2.0
2.63	8.7	43.97	5.3	94.49	1.9
3.01	8.6	45.97	5.2	95.08	1.8
3.43	8.5	47.98	5.1	95.62	1.7
3.89	8.4	50.00	5.0	96.11	1.6
4.38	8.3	52.02	4.9	96.57	1.5
4.92	8.2	54.03	4.8	96.99	1.4
5.51	8.1	56.03	4.7	97.37	1.3
6.14	8.0	58.03	4.6	97.72	1.2
6.81	7.9	59.99	4.5	98.04	1.1
7.55	7.8	61.94	4.4	98.32	1.0
8.33	7.7	63.85	4.3	98.58	.9
9.17	7.6	65.75	4.2	98.82	.8
10.06	7.5	67.48	4.1	99.03	.7
11.03	7.4	69.39	4.0	99.22	.6
12.04	7.3	71.14	3.9	99.39	.5
13.11	7.2	72.85	3.8	99.55	.4
14.25	7.1	74.52	3.7	99.68	.3
15.44	7.0	76.12	3.6	99.80	.2
16.69	6.9	77.68	3.5	99.91	.1
18.01	6.8	79.17	3.4	100.00	.0
19.39	6.7	80.61	3.3		
20.83	6.6	81.99	3.2		

Let us take as an illustrative problem the transmutation to linear scores of the ranks on beauty of the 50 photographs mentioned above. We first find the "per cent position" in the series of ranks occupied by each photograph, by the formula :

$$\text{Per cent position} = \frac{100(R - .5)}{N} \quad . \quad . \quad . \quad (13)$$

In the formula, R is the rank of a particular photograph and N is the total number of photographs in the ranked series. Having this per cent position, we are able to read off directly from Table 63 the corresponding score of amount on a 10-point scale. Thus, for the photograph receiving the rank of 2, we have by the formula :

$$\begin{aligned} \text{Per cent position} &= \frac{100(2 - .5)}{50} \\ &= \frac{150}{50} \\ &= 3 \end{aligned}$$

Looking up the 3 in Table 63, we find that the entry nearest to this value is 3.01, which is seen to correspond to the scale score of 8.6.

In practice it usually saves time to figure out the per cent positions for all the ranks at one time (column 2, Table 64). It will be observed that after this is figured out for the first rank, all that is necessary for the others is to add successively the value $\frac{100}{N}$. In Table 64 this number to be added happens to be a whole number (2), which is easily added. In case this turns out to be a large decimal, it is usually well to figure it out to five or six places and then perform the successive additions on an adding machine. The scale scores can then be looked up in Table 63 all at one time. In case a fractional

TABLE 64

SHOWING THE TRANSMUTATION OF A RANKED SERIES OF 50 ITEMS INTO THE UNITS OF AMOUNT ON AN ORDINARY 10-POINT SCALE

RANK	PER CENT POSITION	SCALE SCORE
1	1	9.3
2	3	8.6
3	5	8.2
4	7	7.9
5	9	7.6
6	11	7.4
7	13	7.2
8	15	7.0
9	17	6.9
10	19	6.7
11	21	6.6
12	23	6.5
13	25	6.4
14	27	6.2
15	29	6.1
16	31	6.0
17	33	5.9
18	35	5.8
19	37	5.7
20	39	5.6
21	41	5.4
22	43	5.3
23	45	5.2
24	47	5.1
25	49	5.0
26	51	5.0
27	53	4.9
28	55	4.8
29	57	4.7
30	59	4.6
31	61	4.4
32	63	4.3
33	65	4.2
34	67	4.1
35	69	4.0
36	71	3.9
37	73	3.8
38	75	3.7
39	77	3.5
40	79	3.4
41	81	3.3
42	83	3.1
43	85	3.0
44	87	2.8
45	89	2.6
46	91	2.4
47	93	2.1
48	95	1.8
49	97	1.4
50	99	.7

rank should be encountered, as occasionally happens, the full formula may be worked out for that particular ranked item.

Since ranked series do not often extend above 50 items and since they are hardly useful when less than 10 items, it has been found rather simple to prepare a table from which may be read off directly the scale equivalents for all ranked series within the above limits. This table, compiled by Mr. Selmar C. Larson, is given in Appendix I (page 491). The table does not give the scale values of fractional ranks. In case one of these fractional ranks should be encountered, it may be given a value midway between the table entries on either side of it.

CONCRETE EXAMPLES OF SCORING CRITERIA BY MEANS OF RANKING

The criterion phase of three aptitude projects will now be described for the purpose of illustrating various uses of the ranking techniques. The first (16) concerns an attempt to devise a battery of aptitude tests for the cashiers in a large department store. In this particular store the cashiers were stationed in special stands. The stands were scattered at frequent intervals throughout the store. The duties of the cashiers were mainly to receive payment for goods sold by the sales girls, to check the bills made out by the latter, and to make change. At some stands the work was heavy; at others, light. The kind of goods surrounding each stand was largely different. The nature of the business made it necessary frequently to change girls from one stand to another. All these things, together with the natural intangibility of merit in this occupation, made it impossible to secure any significant objective criterion. It happened, however, that this store had an unusually able woman in charge of the employment and supervision of these girls. She very

willingly consented to rank the girls on the basis of her intimate observations of their work.

The procedure in ranking was first to write the name of each of the 76 girls on small rectangles of paper about the size of calling cards. After being thoroughly shuffled these cards were given to the supervisor. She first sorted the cards out roughly into three or four grades of ability. Then the cards in the best group were arranged in a series on a table, in the order of merit of the cashiers, the best being placed at the top. Then the next lower group was arranged in a similar manner beneath the first. The same was done with the remaining groups. The series of cards was then reexamined with great care, especially at the junctions of the different preliminary groups, and certain cards were moved up and down in the series until, in the opinion of the superintendent, the cards finally stood in their true order. These cards were then numbered consecutively from 1 to 76. These numbers were taken as the ranks of the respective girls in merit as cashiers. These ranks were then transmuted into units of a linear 10-point scale, which became the criterion score. This example illustrates the technique of securing a criterion score by ranking subjective impressions.

The second example of the use of the ranking technique concerns a project to organize a battery of aptitude tests to predict aptitude in vocal expression (98). This study differs from the one just described in that the behavior judged was of exactly the same kind for all subjects. This was accomplished by taking a special sample of the subjects' behavior for the purpose. This is fairly easy in scholastic aptitudes but usually impossible in industry. Groups of young men ranging in number from 16 to 26 were asked to read as expressively as they could a rather difficult monologue. The men of a given group read the selection one after the other, in the presence of the rest of the group, who were

to judge the merit of the reading. The reader stood back of the judges, so that he was not seen while reading. This was done to eliminate the influence of such irrelevant factors as gesture and personal appearance. Thus each man's performance was judged by each of the others.

The ranking technique is very poorly adapted to the scoring of action criteria. It is accordingly characteristic that in this case, even though a rank criterion was desired, it was necessary first to have the judges tentatively score each reading on a subjective percentage scale as it took place. After all had read, each judge, comparing his impressions of the various performances and with the aid of the previously recorded percentage scores, ranked the various readers in what appeared to him to be their true order of merit. The ranks received by each reader were then averaged,¹ and these averages were themselves ranked. These latter ranks were taken as the final order of merit in vocal expression for that group.

It was necessary to combine the scores of the various groups into a single large group. This was done very simply by transmuting separately each of the ten series of ranks into units of a 10-point scale as described above (page 386). On the assumption that the average and range of ability for the different groups were approximately alike, the resulting linear scores were then closely comparable. The resulting scores of the entire 210 individuals were then massed and treated as a single criterion group, without further regard to the original groups from which they had been assembled. The comparatively high correlations later obtained from these criterion scores indicated that the technique employed must have been practically as effective as where a thoroughly objective criterion is available.

¹ These ranks should really have been converted into linear units before being averaged.

The third and last example of a criterion score secured by the ranking technique concerns free-hand drawing. It will be recalled that this particular investigation is being used at the close of the general account of each step of the aptitude-testing procedure, to illustrate concretely the corresponding stage of an actual investigation.¹ We shall accordingly postpone our account of the securing of this criterion until near the end of the present chapter (page 413), where it will be given in detail.

SCORING CRITERIA BY MEANS OF SUBJECTIVE SCALES

We pass to the consideration of criteria which may be scored by what are metaphorically called "subjective scales." What is really meant by this expression is that when the organism of a competent judge is stimulated by the several degrees of a characteristic type of stimulus pattern there will be evoked a series of symbolic habit reactions, usually in the form of spoken or written numbers. This is what takes place when we judge the width of a sheet of paper to be 9 inches or the weight of a man to be 160 pounds. To be sure, if very much depended upon the accuracy of the determination in the above cases, we should not trust to such a judgment. We should measure the paper with a ruler and weigh the man on a scale. It is fortunate, however, that approximate determinations may be made in this way, because it frequently happens that we desire quantitative information where no possibility of objective measurement exists. While comparatively inexact and varying greatly in precision from person to person, the method of scoring by means of a subjective scale has the advantage of being applicable to a very wide range of criteria otherwise inaccessible to measurement.

Many criteria which might be scored by this method are also susceptible of being ranked. But the ranking method

¹ See pages 300 ff., 325 ff., 359 ff.

is of little value with action criteria, as pointed out above when considering Weaver's criterion for vocal expression. In such cases scoring by a subjective scale is about the only method available. Other cases of this kind are encountered where it is desired to secure a score of the excellence of a particular musical performance or an oratorical effort.

The most conspicuous example of the use of a subjective scale is the traditional method of grading oral recitations, written exercises, and examinations, in the schools. Such scores have been used quite generally as criteria of aptitude in the numerous investigations regarding the validity of the Binet-Simon Tests, the National Intelligence Tests, and the Thorndike College Entrance Tests.

CHARACTERISTIC DEFECTS OF SCORES MADE BY SUBJECTIVE SCALES

One of the most serious difficulties in the use of school marks for criteria appears when the attempt is made to combine the students from a number of different teachers' sections in a particular school subject, so as to make up a group of sufficient size for experimental purposes. The difficulty arises from the fact that, quite apart from imperfect powers of discrimination, different teachers unconsciously mark according to different scales. The differences in the scales employed show themselves by differences in the way their marks run on groups of subjects which, as a whole, presumably are doing about the same quality of work.

Subjective marking scales may produce results which differ in two important respects. First, a given teacher or judge may tend to mark all work generally high or generally low. This tendency is revealed by marked differences in the arithmetical averages of comparable groups of subjects scored by different people. Secondly, a given judge may hold his marks closely around his mean, whereas another may give

marks ranging widely from his mean. This second tendency is revealed by the standard deviation of comparable groups being markedly different. Discrepancies between the averages of marks and other scores from comparable groups made by different judges on the basis of subjective scales are usually recognized. But differences in the dispersion or range of marks, as a rule, receive little or no attention. Yet the second type of inequality in marking is fully as important as the former. One of several difficulties from this source is encountered when the averages of several sets of scores are being made for a single joint criterion. For example, if two sets of grades or scores are being averaged for a series of individuals and the first has twice as large a standard deviation as the second, the first will have twice as much influence on the average as the second. It is as if the first were given double weight in making the average. Table 65 from a freshman class at the University of Wisconsin shows one or two striking differences of both the kinds mentioned above.

TABLE 65

SHOWING THE DIFFERENCES IN THE MEANS AND THE SPREADS OF GRADES
IN VARIOUS COURSES OF THE FIRST SEMESTER OF AN ENGINEERING
COLLEGE

COURSE	MEAN	STANDARD DEVIATION
Chemistry	81.9	10.2
Drawing	81.5	9.2
English	78.1	8.8
Mathematics	80.9	9.9
Shop I	85.9	4.0
Shop IV	80.0	10.0

There is a range of nearly 8 points in the means, and the greatest spread or dispersion is over $2\frac{1}{2}$ times that of the least.

It is accordingly not wise to accept uncritically the criterion scores from several small groups where each group has been scored by the subjective scale of a different person, if all are to be combined into a single trial group for experimental purposes. A typical case in point is encountered when it is desired to combine several sections of a given school subject in order to make up a sufficiently large experimental group, each section being under a different instructor. We have in such a situation not only the peculiarities of each instructor's marking, but more or less extensive differences in the instruction itself. Owing to this latter factor alone there might result, even with the most objective criterion, a marked difference between the scores of the different groups which would be in no way related to any differences in aptitude.

METHOD OF RENDERING COMPARABLE, SUBJECTIVE
SCORES BY DIFFERENT JUDGES

Fortunately there is a method by which the most diverse sets of scores may be made sufficiently comparable for the practical purpose of combining small groups of subjects into a single large trial group. The method, in brief, is to convert the various series of non-comparable scores into new series which shall be comparable (38). To be comparable, series of scores representing groups drawn at random from the same population should have approximately the same mean and the same standard deviation. Accordingly, the method of securing comparability is to convert the non-comparable series into series which shall have the same means and the same standard deviations. If there is substantial reason to believe that any of the sub-groups are in some sense selected and differ materially in range and amount of aptitude from the others, such groups should, of course, be rejected. It should be added, also, that the larger the sub-groups the more adequate will the conversion be. Groups of subjects

smaller than about 20 should not be included if sufficient subjects can be secured from larger groups.

In converting scores into comparable series the first thing to do is to calculate the means and standard deviations of the original series. In case one of the original series, especially if it be a large one, chances to have both the mean and the standard deviation desired for the criterion, this series may be taken as the standard and the remaining series converted into new ones which are comparable to it. If, however, none of the original series has the statistical characteristics required of the criterion, then all may be converted into new series which shall have any mean and standard deviation desired.

The method of conversion (38) is as follows :

Let M = the mean of the original series.

Let σ = the standard deviation of the original series.

Let X = a given individual's score in the original series.

Also,

Let M' = the mean of the new series.

Let σ' = the standard deviation of the new series.

Let X' = the same individual's score in the new series. This is to be found.

Then,

$$X' = K + SX \quad . \quad . \quad . \quad . \quad . \quad (14)$$

where

$$S = \frac{\sigma'}{\sigma} \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

and
$$K = M' - SM \quad . \quad . \quad . \quad . \quad . \quad (16)$$

In case S comes out a decimal (as it usually does), the decimal should be carried out to four or five places.

The method may be illustrated by a conversion based on the following data: Suppose that a series of scores was found by actual computation to have a mean of 30 and a σ

of 3.5. One man in the group had a score of 27. What would this man's score be if converted into the equivalent school mark? It is known that an ideal series of school marks has a mean of about 81 and a standard deviation of about 7. Then our various values are as follows:

ORIGINAL SERIES

$$M = 30$$

$$\sigma = 3.5$$

$$X = 27$$

NEW SERIES

$$M' = 81$$

$$\sigma' = 7$$

$$X' \text{ is to be found}$$

Accordingly,

$$S = \frac{7}{3.5} = 2$$

$$K = 81 - 2 \times 30 = 21$$

Therefore by formula,

$$\begin{aligned} X' &= 21 + 2 \times 27 \\ &= 75 \end{aligned}$$

The use of the method in practice, where entire series are involved, may be illustrated by converting into comparable scores the marks of three sections of beginning pupils in high-school French. The original scores are given in the first three columns of Table 66; those after conversion to comparable series, in the last three columns.

In cases like those shown in Table 66, where any considerable number of scores are to be transmuted, the labor may be reduced to very small proportions by the construction of a systematic table from which the equivalent values later may be read off directly. With the aid of a good calculating machine like the Monroe, or even an adding machine, this table may be constructed in five or ten minutes. K is placed on the dial from which sums are read, and S is placed on the keyboard. The machine is then caused to add S continuously, one addition being made for every value of X from 1 to the maximum value in the series. A reading (X') is taken

TABLE 66

SHOWING THE SEMESTER MARKS IN BEGINNING FRENCH AS GIVEN BY THREE DIFFERENT TEACHERS TO AS MANY DIFFERENT SECTIONS IN THE SAME SCHOOL, AND THE SAME AS TRANSMUTED SO THAT ALL SECTIONS HAVE THE SAME MEAN AND THE SAME STANDARD DEVIATION

ORIGINAL MARKS IN BEGINNING FRENCH				MARKS IN BEGINNING FRENCH TRANSMUTED SO THAT ALL SECTIONS HAVE THE SAME MEAN AND THE SAME STANDARD DEVIATION		
Section I (Teacher Y.)	Section II (Teacher K.)	Section III (Teacher S.)	Section I (Teacher Y.)	Section II (Teacher K.)	Section III (Teacher S.)	
90	50	82	90.5	63.1	80.8	
52	79	87	61.0	81.1	85.0	
70	94	87	75.0	90.8	85.0	
71	69	95	74.2	75.1	91.9	
72	86	66	76.5	85.8	67.2	
82	94	84	84.3	90.8	82.5	
95	70	90	94.4	75.7	87.6	
79	82	93	82.0	83.3	90.1	
85	50	59	86.6	63.1	61.2	
70	90	76	75.0	88.3	75.7	
70	68	70	75.0	74.5	74.8	
87	70	85	88.2	75.7	83.3	
91	93	98	91.3	90.2	94.4	
79	50	66	82.0	63.1	67.2	
72	75	96	76.5	78.9	92.7	
78	73	94	81.2	77.6	91.0	
52	93	81	61.0	90.2	79.9	
88	79	75	89.0	81.4	74.8	
95	75	75	94.4	78.9	74.8	
80	94	81	82.7	90.8	79.9	
75	69		78.9	75.1		
	96			92.1		
	91			89.0		
	91			89.0		
Mean 77.76	78.88	82.25	80.94	80.98	80.99	
S.D. 11.576	14.283	10.57	9.04	8.99	8.999	

at each addition within the range of values found in the data to be transmuted. For example, in the first column of Table 66, $M = 77.76$ and $\sigma = 11.58$. Accordingly, $S = .77772$ and $K = 20.565$. Putting these S and K values on the machine as described above, we have in a few moments a table of equivalent values such as is shown in Table 67.

TABLE 67

A TYPICAL TABLE FOR TRANSMUTING A SERIES OF SCORES INTO A NEW SERIES HAVING ANY DESIRED MEAN AND STANDARD DEVIATION

In this case it was desired to have the new series possess a mean of 81 and a standard deviation of 9. This table was used to transmute Section I of Table 66.

ORIGINAL SCORE (X)	TRANSMUTED SCORE (X')	ORIGINAL SCORE (X)	TRANSMUTED SCORE (X')
50	59.4	75	78.9
51	60.2	76	79.6
52	61.0	77	80.4
53	61.8	78	81.2
54	62.5	79	82.0
55	63.3	80	82.7
56	64.1	81	83.5
57	64.9	82	84.3
58	65.6	83	85.1
59	66.4	84	85.8
60	67.2	85	86.6
61	68.0	86	87.4
62	68.8	87	88.2
63	69.5	88	89.0
64	70.3	89	89.7
65	71.1	90	90.5
66	71.9	91	91.3
67	72.6	92	92.1
68	73.4	93	92.8
69	74.2	94	93.6
70	75.0	95	94.4
71	75.7	96	95.2
72	76.5	97	96.0
73	77.3	98	96.7
74	78.1	99	97.5

RATING SCALES

While the ordinary academic marking system represents perhaps the most widely used subjective scale, it is by no means the only one. In cases where the criterion and the scale are both subjective, it is frequently felt that a scale consisting of only five or seven steps is preferable to the comparatively minute gradations of the academic marking system. On the other hand, excessively coarse scales of only three steps have been suggested and used, though they are of doubtful value for criterion purposes. These simplified subjective scales are usually called *rating scales*. In their most primitive form they are employed in about the same way as is the academic marking system. The principal of a school may, for example, be asked to score his grade teachers on their general efficiency as teachers, using a 5-point scale. He will simply sit down at his desk and, with the names of the teachers before him, write down a number opposite each. If a teacher is regarded as of the highest efficiency, he will write down opposite her name a 5, if of average efficiency a 3, if of the lowest efficiency a 1, and so on.

It has frequently been urged against numerical scales that it is difficult to compare one's impressions of a person, say, with a number. This form of statement probably exaggerates the difficulty, though it is doubtless true that inexperienced persons do have trouble in assigning numerical values to the several degrees of certain variables. The reason is obvious. The particular verbal habits required in this case have not yet been established in those persons. The natural solution of the difficulty has been to devise scales based on habits already well established. A very common form of non-numerical scale is one made up of a series of adjectives which indicate different degrees of a trait. A common five-step scale of this kind is:

Poor, fair, average, good, excellent

Another of the same general form is :

Lowest, low, middle, high, highest

Probably a more effective form of this type of scale is that in which the steps of the scale are marked, not by more or less colorless adjectives as above, but by descriptive expressions which apply specifically to the trait being rated. For example, if one were rating the ability of a person to learn a complicated trade, it would probably be more effective to use such terms as :

Stupid, dull, average, clever, brilliant

since most people have fairly well-established habits of applying these terms to learning behavior. Or if one were rating a person in respect to industriousness, the following series might be used :

Lazy, easy-going, steady, brisk, speedy

As a movement in the direction of a still greater specificity in the subjective or habit scale, there have been constructed scales in which pungent descriptions of actual behavior or results of behavior replace the single words of the adjective scales. Thus the adjective scale for ease of learning might appear as follows :

Has to be told over and over	Requires frequent explanation	Learns with- out much trouble	Picks it up very quickly	Hardly needs to be told
------------------------------------	-------------------------------------	-------------------------------------	--------------------------------	-------------------------------

Similarly the scale for industriousness might appear as :

Idles unless watched	Requires frequent prodding	Works fairly steadily	Brisk and active	Works at top speed
----------------------------	----------------------------------	-----------------------------	------------------------	--------------------------

THE MAN-TO-MAN RATING SCALE

The difficulty experienced by certain persons in "comparing a man with a number" gave rise to the suggestion that the best possible solution in such cases would be to compare a man with another man. This consideration led Walter Dill Scott during the late war to construct his famous man-to-man rating scale. This scale was designed for use by army officers chiefly in rating subordinates. It was hoped through its use to secure promotions in the service more nearly on the basis of true merit.

The system as finally worked out involved a 5-point scale which was unique in that it was constructed in the main by the rater himself. If an officer were to rate a number of subordinates on their general value to the service, he would be given a blank form such as :

Highest	40
High	32
Middle	24
Low	16
Lowest	8

He would then be directed to select from officers ten or more men with whom he has served or with whom he is well acquainted. The names of these men would be written on slips of paper. These slips would be arranged, in order of rank in general value of the men to the service, from highest to lowest. He would then pick the highest man of the series and enter his name on the first line opposite the number 40. He would next place the lowest name on the lowest line opposite 8. From halfway between the extremes of his series he would take a third name which would be recorded on the middle line opposite 24. Next he would select a name midway between middle and highest, placing it on the second

line, and a name midway between middle and lowest, placing it on the fourth line. The scale would then appear as follows :

Highest	<i>Betts</i>	40
High	<i>Knight</i>	32
Middle	<i>Hamilton</i>	24
Low	<i>Aldrich</i>	16
Lowest	<i>Newton</i>	8

In use, the rater is expected to compare the man he is rating, as regards general value to the service, with the men on his rating scale (Betts, Knight, etc.). When this has been done, the man being rated will receive as a score the number opposite the name of the man of the scale he most nearly resembles in this particular trait. For example, if Captain Shields is judging Lieutenant Murphy and decides that with time the lieutenant will mature into an officer as good as Knight, Murphy will receive a score of 32.

One of the weaknesses of a man-to-man scale in aptitude-criterion work is that it must be constructed by the person who is to use it. In this respect such scales differ sharply from ordinary rating scales, which usually are constructed by persons who have had special training for the work. This constitutes a rather serious defect, because many fairly competent judges would hardly have the ability without special training to perform the somewhat complicated task of constructing the scale. Still others would lack the patience. If such a scale is to be used in securing a criterion score, the rater should be carefully supervised by the aptitude psychologist.

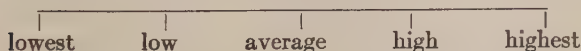
GRAPHIC RATING SCALES

An important variant of the rating scales as described above consists in having the scale printed on a blank form in

such a way that the rater can indicate his rating by merely checking with a pencil that step of a scale upon which a person rated is believed to fall. This is especially common where it is desired to secure separate ratings on a variety of aspects of a vocation. In most cases these separate ratings later would be combined into a single criterion score.

The checking blank just described is a transitional stage leading to an important type of rating device known as the *graphic rating scale*. The novel feature of the graphic rating scale is that the range of the trait being scored is represented by a straight line, usually three or four inches in length. In some way, usually by means of descriptive expressions placed beneath, the various steps of the scale are indicated. When scoring the subjective impression of a person on such a scale, the scorer glances along the line, noting the various descriptive expressions beneath. He then places a check mark on the line at the point where the person being scored seems to fall. Since the check mark may be placed anywhere on the line, the method permits as fine distinctions as the judge can make. Numerical scores may be obtained by constructing a scale of some convenient material, such as cardboard, on the edge of which a distance equal to that of the checking line is divided off into 10 or 20 units, with appropriate numbers added. This scale can then be placed near the line, and the numerical equivalent of the position of the check may be read off to the nearest unit.

Just as there are three general types of ordinary rating scales, so three general types of expressions may be placed beneath the line marking the steps of the scale. First there may be placed a series of colorless adjectives such as :



Secondly, there may be used a series of graded expressions which apply specifically to the particular trait being rated. For example, a graphic rating scale for elementary teachers in respect to ability to get on with children might be :

Children will do anything for her	Occasional fric- tion with dif- ficult pupils	Children dis- like her thoroughly
---	---	---

A third method of indicating the steps on a graphic rating scale to be used for scoring the subjective impressions of persons is to have written beneath the line the names of individuals representing the various degrees of the trait. In such cases, of course, the individuals whose names appear beneath the line must be known to the rater. The method of constructing such a series of names has been described above (page 403) in connection with the man-to-man rating scale.

A rather elaborate graphic rating scale using both adjectives and persons' names to mark the intervals was used by Edward M. Martin and Herbert A. Toops (56) in securing criterion ratings on the patrolmen of a police force. A mimeographed sheet, shown in Figure 55, furnished the basis of the system. Accompanying this was a cardboard stencil, shown in the stippled portion, and a cardboard scale. The stencil contains first a conventional 5-step adjective scale. Immediately above each adjective, however, is a blank space of similar size in which the police officer doing the scoring is to write the names of the patrolmen who, in all of his experience, fall nearest the various points of the scale. These names are to be chosen by the general method described for the army rating scale. Thus each scoring officer must have his own private scale. This is the reason for having a stencil separate from the mimeographed sheet. Each rater must have his own stencil, but the same mimeo-

SERVICE RATING SHEET

Name Guthrie Precinct 4 Date 5/7/25

I. Appearance

Physique: Athletic or corpulent
Neatness: Consider person and dress
Bearing: Military attitude and carriage

Lowest ☒ Low ☒ Middle ☒ High ☒ Highest ☒

II. Intelligence

Ability to write clear and legible report
Does he act with good judgment without instructions?
Does he answer questions intelligently?

Lowest ☒ Low ☒ Middle ☒ High ☒ Highest ☒

III. Discipline

Is he punctual?
Is he respectful to commanding officers?
Does he obey orders promptly and cheerfully?

Lowest ☒ Low ☒ Middle ☒ High ☒ Highest ☒

IV. Efficiency

Does he keep his beat in good condition without arousing ill-feeling among the residents?
Does he "keep his head" in an emergency?
Is he courteous to the public?
Does he notice violations of ordinances?

Lowest ☒ Low ☒ Middle ☒ High ☒ Highest ☒

Rating Officer (Initials) A. V. S.

FIG. 55. Martin's graphic rating scale for policemen. The stippled portion is a cardboard mask placed over the mimeographed sheet to serve as a guide to the rater in placing his check marks on the various lines. (See Figure 56.)

graphed sheet will do for all. In use, this stencil is placed over the mimeographed sheet in such a way that the lines of the sheet to be checked will show through the rectangular holes, as illustrated in Figure 55. A check mark is then made on each of the lines representing the scales of the four traits being scored. The stencil is then removed from the sheet and the cardboard scale placed next to each line in succession. In the stippled part of Figure 56 this cardboard scale is shown in position to measure the rating of Patrolman Guthrie as judged by Officer Stossel. The score in this instance is clearly seen to be 15 on a scale of 20.

IDEAL DISTRIBUTION OF POPULATION ON RATING SCALES

It should be remembered that for the adjectives and other descriptive devices to represent a true linear scale, the population being rated should not be distributed equally under each of the five divisions of the scale. This will readily be understood by referring once more to Figure 54. In round numbers the percentages of the statistical population normally falling under each of the adjectives of a 5-point scale are as follows:

TABLE 68

IDEAL DISTRIBUTION OF POPULATION ON A 5-POINT SCALE

NUMERICAL VALUE	ADJECTIVE SCALE	PER CENT OF POPULATION FALLING WITHIN EACH INTERVAL OF THE SCALE
5	Excellent	6%
4	Good	24%
3	Average	40%
2	Fair	24%
1	Poor	6%

The reader will understand, of course, that the above percentages represent a marked tendency and are not absolute.

SERVICE RATING SHEET

Name Guthrie Precinct 4 Date 7/7/25

I. Appearance

Physique: Athletic or corpulent
Neatness: Consider person and dress
Bearing: Military attitude and carriage

II. Intelligence

Ability to write clear and legible report
Does he act with good judgment without instructions?
Does he answer questions intelligently?



III. Discipline

Is he punctual?
Is he respectful to commanding officers?
Does he obey orders promptly and cheerfully?

IV. Efficiency

Does he keep his beat in good condition without arousing ill-feeling among the residents?
Does he "keep his head" in an emergency?
Is he courteous to the public?
Does he notice violations of ordinances?

Rating Officer (Initials) R. T. L.

FIG. 56. Martin's graphic rating scale for policemen. The stippled portion in this figure shows a cardboard scale of twenty steps for securing a numerical measure of the rating after it has been made.

But since it is the most probable distribution of even a moderate number of data, the person doing the scoring should be instructed carefully in the usual distribution on the scale.

The trouble with the suggestion just made is that ordina-

rily it will be impossible to tell whether or not a given judge has followed the principle. The reason for this is that if he were scoring enough individuals for the percentages to appear with any clearness, there would be enough subjects to rank, in which case the criterion could (and probably should) be treated by the method of ranks as described above (page 386). This does not apply, however, to action criteria or to product criteria which cannot readily be moved about, in which the ranking technique has little value. This means that in the case of most product criteria and all subjective impression criteria the use of subjective scales of any kind will ordinarily be limited to situations where fewer than a dozen or twenty items are being scored by a single judge. And here, even more than by the ranking method, it is desirable that as many independent scores should be obtained from as many different judges as possible, all being averaged together for the final criterion score.

QUALITATIVE SCALES MADE FROM SERIES OF SPECIMENS

The fifth and last method of scoring criteria which we shall consider is that performed by means of qualitative or specimen scales. Perhaps the best-known scales of this type are those developed for measuring the quality of handwriting. Such scales are useful where the thing to be measured is objective, yet not measurable by the ordinary scales known to physical science. Examples of criterion products measurable by such scales are: the quality of English composition, the accuracy of a free-hand representative drawing, the excellence of a job of soldering, and the "finish" of a piece of work turned out on an engine lathe. Such scales are made up of a number of specimens or samples of the product to be measured, arranged in a series convenient for handling. These scale specimens must, of course, be known

14

He has achieved Success laugh
often and loved much who has
gained the respect of intelligent
men and the love of little

15

He has achieved success who has lived
well, laugh laugh often and love much,
who has gained the respect of intelligent men
and the love of little children, who has

16

He has achieved success who has
lived well, laugh often and love
much, who has gained the respect
of intelligent men and the love of

17

He has achieved who has
well, laughed often and loved
who has gained the respect of
intelligent man and the love.

FIG. 57. Four steps of the Starch Handwriting Scale, an example of a qualitative scale made up from a series of specimens. (Reproduced by permission of the author.)

by some indirect means to be separated by equal amounts of whatever is being measured.

In this, a specimen scale differs from an ordinary scale such as a foot rule. For example, of the four samples of handwriting shown in Figure 57, No. 15 is as much better

in general merit than No. 14 as No. 16 is better than No. 15. Moreover, the difference in general merit between No. 14 and No. 16 is twice as great as the difference between No. 14 and No. 15. It is as if we had, instead of an 18-inch rule, 18 separate sticks arranged in a series, the smallest one inch in length and each of the others one inch longer than the one next to it. In this latter case, just as in the case of the handwriting samples of Figure 57, the *difference* between sticks 14 and 15 would be the same as the *difference* between sticks 15 and 16; and likewise the difference between sticks 14 and 16 would be twice as great as that between sticks 14 and 15.

In constructing specimen scales the method usually employed is first to have a large number of random samples of whatever is to be measured by the scale, ranked by a large number of competent judges. From these rankings there are derived by statistical procedures quantitative or scale values for each specimen. This process is called *scaling*. If the number of samples ranked is large and they are scattered over the range to be measured in a statistically normal manner, it is usually possible to find among the scaled samples specimens falling close to each of the points marking off the even units of whatever scale is adopted. These samples thus fortunately located are chosen to make up the specimen scale, the remaining ones being discarded. Thus there will appear on the finished scale that specimen which when scaled was found to be nearest to an even 8 units of merit, that specimen which was found to be nearest to an even 9 units of merit, and so on throughout the range of the scale.

The specimens thus carefully chosen are then arranged, usually side by side, according to their scale values and in some portable form. When it is desired to measure a sample of unknown value by means of the scale, the sample is moved

along the scale series until a point is reached at which the scale-specimen appears most to resemble the unknown in the magnitude of whatever is being measured. The numerical value of this scale-specimen is then taken as the amount of merit possessed by the sample being measured.

A CRITERION IN DETAIL — FREE-HAND DRAWING

At this point it will be profitable to turn from the discussion of criteria in general to the detailed consideration of a concrete example. As in the chapters which have been devoted to the earlier stages of the aptitude project, we shall choose for this purpose free-hand drawing, thus advancing the detailed consideration of our type project another step.

It has already been pointed out that before beginning the testing work on this project, the possibility of securing a satisfactory criterion was assured. The plan was to have the subjects, at the close of the course, make careful representative drawings of some fairly complicated object. To this end, the subjects were taken in groups to a room in a local historical museum where there was on exhibition an antique melodion. The subjects were seated at about equal distances from the instrument, given uniform sheets of drawing paper, and directed to make as perfect perspective drawings of the melodion as possible. Special instructions were given that the drawings were not to be "artistic" but should be realistic and purely representative. The drawings were to be of a certain size, and a limit of one hour was placed on the amount of time available to complete the work. The subjects were warned from time to time of the progress of the interval so that they might adopt a speed which would complete the drawing in the specified period. As the interval drew toward its close, the slow subjects were warned and urged to complete their drawings as best they could. At the

conclusion of the period the subjects wrote their names on the backs of the drawings and handed them in. One of the better drawings thus produced, that of subject No. 1, is shown in Figure 58.

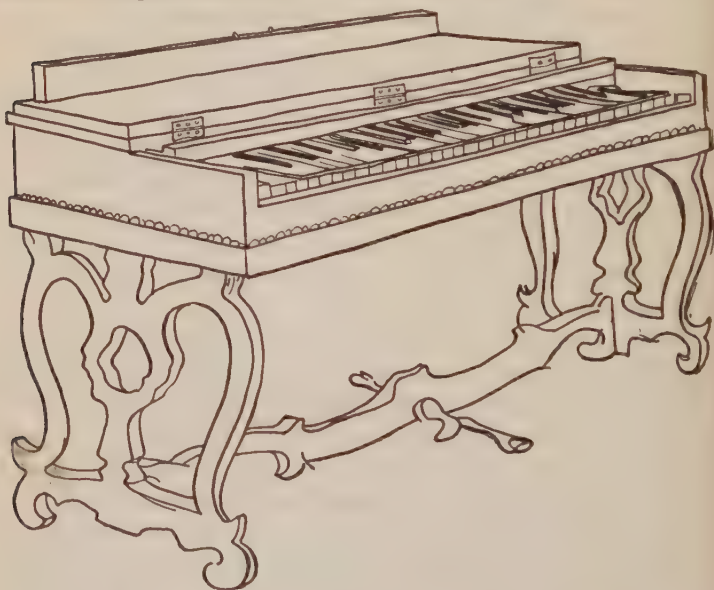


FIG. 58. The criterion performance of Alcott, subject No. 1.

Such drawings, while clearly product criteria, evidently are not measurable in terms of any ordinary objective scale. They may easily be ranked, however. Accordingly, four members of the university faculty who were experts in judging drawing were asked independently to arrange the drawings in their order of general merit. Each expert first examined carefully the instrument represented in the drawings and was informed of the instructions given the subjects. A given judge would then sort over the drawings on a long table, first arranging them roughly into three piles on the

basis of merit. Then the drawings of the best pile were arranged side by side in their order of general excellence. The drawings of the next best pile were next added to the series in the order of descending merit, and so on. At the end of the process there were 48 drawings arranged side by side in a descending order from left to right. The expert then went over the series several times, revising his arrangement here and there by moving one drawing up and another down the series, until it seemed to him a perfect arrangement. The order of the various drawings was then recorded by the experimenter. The drawings were again carefully shuffled and the procedure repeated with the remaining judges, one at a time. All the judges were, of course, kept in complete ignorance of the rankings made by the others.

The four sets of ranks thus secured are shown in Table 69. It will be noted in this table that, as usual in such cases, the subjects are arranged in alphabetical order to facilitate the location of the individuals by the experimenter while recording data. As a result the ranks follow no particular sequence as one goes down the column. It may be observed by comparing the ranks given to any particular drawing by several judges that, while there was some difference of opinion, the tendency to agreement is very marked. It is at once evident, for example, that Alcott's drawing is very generally regarded as superior whereas that of Bailey is quite as generally regarded as inferior.

The reader is already familiar with the fact (page 383) that ranks, as such, cannot be treated like ordinary scale units. The four sets of ranks of Table 69 were accordingly converted directly into equivalent linear scores by the use of the conversion table (Appendix I, page 491). The converted scores thus obtained are shown in the first four columns of Table 70.

The criterion data are now in a form which permit of manipulation by ordinary arithmetical methods. They

TABLE 69

SHOWING THE RANKS ASSIGNED THE DRAWINGS OF THE VARIOUS
SUBJECTS BY THE DIFFERENT JUDGES (Strader)

SUBJECT	RANKS BY JUDGE 1	RANKS BY JUDGE 2	RANKS BY JUDGE 3	RANKS BY JUDGE 4
Alcott	2	7	2	6
Bailey	43	39	44	43
Beatty	15	3	8	4
Bemis	36	29	34	21
Brayton	38	21	29	33
Cunningham	7	8	15	3
De Vries	5	10	4	8
Drake	9	2	11	2
Folsom	21	25	20	20
Forrest	23	37	16	32
Frankfurth	28	28	21	27
Gates	25	22	28	35
Guenther	8	12	9	5
Hagberg	4	35	12	30
Hermann	32	42	25	40
Hiestand	16	5	13	10
Huntzicker	14	4	1	7
Jensen	45	44	46	38
Johnson, M.	47	48	47	48
Johnson, R.	33	15	35	16
Kampen	40	45	22	44
King	48	36	48	46
Klint	13	19	14	15
Krebs	24	23	6	14
Krueger	17	18	31	24
Kuehn	46	46	40	47
Lauter	31	40	39	39
Leadstone	22	9	33	17
Mathias	1	1	3	1
Monk	39	43	36	42
McCullough	30	17	43	29
Mueller	19	20	27	22
Olson	12	13	7	18
Patterson	6	14	5	9
Portman	26	31	18	28
Purcell, C.	18	41	23	23
Purcell, K.	44	47	42	36
Rex	42	32	41	45
Ruste	27	38	38	31
Schmitz	41	33	45	41
Stillman	11	6	17	13
Thomas	29	30	19	37
Thorbus	10	11	32	11
Trafford	20	27	30	25
Vaughan	34	24	37	26
Vivian	37	34	26	34
Welter	35	26	24	19
Worst	3	16	10	12

TABLE 70
RANKS TRANSMUTED INTO UNITS OF AMOUNT

SUBJECT	JUDGE 1	JUDGE 2	JUDGE 3	JUDGE 4	JUDGES 1 AND 2	JUDGES 3 AND 4	CRITERION SCORE, SUM OF ALL FOUR JUDGES' RATINGS	FINAL TRANS- MUTED CRITE- RION SCORE IN TERMS OF SCHOOL MARKS
Alcott . . .	8.6	7.4	8.6	7.2	16.0	15.8	31.8	93
Bailey . . .	2.6	2.6	2.4	3.3	5.2	5.7	10.9	72
Beatty . . .	6.0	7.8	7.0	8.2	13.8	15.2	29.0	90
Bemis . . .	3.7	5.4	4.0	4.5	9.1	8.5	17.6	79
Brayton . .	3.5	4.1	4.5	5.4	7.6	9.9	17.5	79
Cunningham	7.2	8.2	6.0	7.0	15.4	13.0	28.4	89
De Vries . .	7.6	7.0	7.8	6.7	14.6	14.5	29.1	90
Drake . . .	6.8	8.6	6.5	8.6	15.4	15.1	30.5	91
Folsom . . .	5.4	5.5	5.5	4.9	10.9	10.4	21.3	82
Forrest . . .	5.2	4.2	5.9	3.6	9.4	9.5	18.9	80
Frankfurth .	4.6	4.7	5.4	4.6	9.3	10.0	19.3	80
Gates . . .	4.9	3.9	4.6	5.3	8.8	9.9	18.7	80
Guenther . .	7.0	7.6	6.8	6.4	14.6	13.2	27.8	89
Hagberg . . .	7.8	4.4	6.4	3.9	12.2	10.3	22.5	83
Hermann . .	4.2	3.2	4.9	2.8	7.4	7.7	15.1	76
Hiestand . .	5.9	6.7	6.3	7.6	12.6	13.9	26.5	87
Huntzicker .	6.1	7.2	9.3	7.8	13.3	17.1	30.4	91
Jensen . . .	2.2	3.5	1.9	2.4	5.7	4.3	10.0	71
Johnson, M.	1.4	.7	1.4	.7	2.1	2.1	4.2	65
Johnson, R.	4.1	5.9	3.9	6.0	10.0	9.9	19.9	81
Kampen . . .	3.3	2.4	5.3	2.2	5.7	7.5	13.2	74
King	.7	1.9	.7	3.7	2.6	4.4	7.0	68
Klint	6.3	6.0	6.1	5.6	12.3	11.7	24.0	85
Krebs	5.1	6.1	7.4	5.2	11.2	12.6	23.8	85
Krueger . . .	5.8	5.1	4.3	5.7	10.9	10.0	20.9	82
Kuehn	1.9	1.4	3.2	1.9	3.3	5.1	8.4	69
Lauter	4.3	3.3	3.3	3.2	7.6	6.5	14.1	75
Leadstone . .	5.3	5.8	4.1	6.8	11.1	10.9	22.0	83
Mathias . . .	9.3	9.3	8.2	9.3	18.6	17.5	36.1	97
Monk	3.2	2.8	3.7	2.6	6.0	6.3	12.3	73
McCullough .	4.4	4.5	2.6	5.8	8.9	8.4	17.3	78
Mueller . . .	5.6	5.3	4.7	5.5	10.9	10.2	21.1	82
Olson	6.4	5.7	7.2	6.3	12.1	13.5	25.6	87
Patterson . .	7.4	6.8	7.6	6.1	14.2	13.7	27.9	89
Portman . . .	4.8	4.6	5.7	4.3	9.4	10.0	19.4	80
Purcell, C. .	5.7	5.2	5.2	3.0	10.9	8.2	19.1	80
Purcell, K. .	2.4	3.7	2.8	1.4	6.1	4.2	10.3	71
Rex	2.8	2.2	3.0	4.2	5.0	7.2	12.2	73
Ruste	4.7	4.3	3.5	3.5	9.0	7.0	16.0	77
Schmitz . . .	3.0	3.0	2.2	4.1	6.0	6.3	12.3	73
Stillman . . .	6.5	6.3	5.8	7.4	12.8	13.2	26.0	78
Thomas	4.5	3.6	5.6	4.4	8.1	10.0	18.1	79
Thorbus . . .	6.7	6.5	4.2	6.5	13.2	10.7	23.9	85
Trafford . . .	5.5	4.9	4.4	4.7	10.4	9.1	19.5	80
Vaughan . . .	4.0	4.8	3.6	5.1	8.8	8.7	17.5	78
Vivian	3.6	4.0	4.8	4.0	7.6	8.8	16.4	77
Welter	3.9	5.6	5.1	4.8	9.5	9.9	19.4	80
Worst	8.2	6.4	6.7	5.9	14.6	12.6	27.2	88

may accordingly be combined to make up a single criterion score which will be a more accurate measure of the true excellence of the drawings than any one judge's ratings alone. This has been done by simple addition. The results appear in column No. 7 of Table 70. This is the criterion score for the present project.

Since we had four independent measures of the drawings which constituted our criterion, it was possible to secure an excellent coefficient of reliability. This was done by combining the ratings of the first two judges for one measure of the drawings and combining the ratings of the second two judges for a second measure. The correlation between the two measures thus secured was then computed. It was found to be $+.924$. This represents the reliability coefficient for two judges. The reliability for all four judges may be found by the Spearman-Brown formula :

$$\frac{2r}{1+r}$$

Substituting appropriately in this formula, we obtain the value of r , $+.96$, as the reliability coefficient of the ratings of the drawings. Extracting the square root of this in turn, we have $+.98$ as the approximate correlation between the total or average of these four judges' ratings and the true measure of the merit of the series of drawings. These results indicate that even with such seemingly intangible media as merit of free-hand drawings, surprisingly accurate determinations may be made by the careful use of an appropriate technique.

The reader must not make the mistake of assuming that the above reliability coefficient of $+.96$ is the reliability of the criterion score. As already stated, it is only the reliability of the measurement of the criterion *product*. A *real* reliability coefficient for the criterion might have been secured by re-

quiring the subject to make, on a subsequent day, a drawing of an object very similar to the one already described. After these second drawings had been scored and the four judges' ratings combined, a true reliability coefficient could have been secured by correlating this score with the four combined ratings of the first set of drawings. This would certainly have been much lower than .924, due to the fact of accidental personal variation from day to day, and so on. As a matter of fact a second drawing was secured, but as it was a highly shaded one of a large plaster scroll, it was hardly comparable with the drawing of the melodion. The reliability of the ratings on this second drawing was +.956, or almost exactly that of the first. When the combined ratings of the two sets of drawings were correlated, the result was only +.638. Owing to the obvious differences between the two sets of drawings, this value of .638 is doubtless an understatement of the reliability of the criterion actually used.

There remains one more detail before the criterion is ready for use. It will be observed that as it stands in Table 70, the scores range from 4 to 36. This is a scale of merit with which no one is familiar. It would consequently greatly limit the usefulness of a battery if the regression equation were to be constructed so as to forecast drawing aptitude on such an unfamiliar scale. Fortunately it is a simple matter to transmute the scores of column 7 into the equivalents of school grades by Formula 14 (page 397). First, the mean was found to be 20.008 and the standard deviation to be 7.033. If we say that the equivalent series of school marks should have a mean of about 81 with a σ of 7, then the S of the formula becomes $\frac{7}{7.033}$, or .996, and the K becomes $81 - .996 \times 20.01$, or 61.09. The conversion formula in this case then becomes:

$$X' = 61.09 + .996 X$$

Applying this formula to the data of column No. 7 of Table 70, and recording the results to the nearest two-place number (to reduce labor in subsequent computations), we have column No. 8, our final criterion score.

CHAPTER THIRTEEN

SELECTING THE FINAL APTITUDE BATTERY

IN the preceding pages we have traced the process by which were secured from the trial subjects: first, a complete set of scores on the tests of the preliminary battery, and second, a complete set of the all-important criterion scores. With these two sets of scores available it now becomes possible to determine which of the tests of the preliminary series are worthy to be retained. Those tests which are found by this process to be without value are at once discarded. This is the final stage of the procedure often spoken of as "testing the tests."

In the last analysis, whether a test will be retained in the final battery is dependent upon whether or not it contributes enough to the prognostic efficiency of the battery as a whole to repay the cost incidental to its use. A cheap test quickly given and easily scored might be retained even if contributing relatively little to the yield of the battery, whereas an expensive test, requiring much time to score, might be rejected from the same battery even if contributing more to the yield than the former. Moreover, the contribution of a test to the forecasting efficiency of a battery is dependent jointly upon the nature and extent of the relation between the test and the criterion on the one hand, and upon the nature and extent of the relation among the tests themselves on the other (pages 254 ff., 271 ff., 450 ff.). We shall not, at present, trouble ourselves about the details of these various relations, but shall proceed at once to the consideration of the standard measure of such relations — the correlation coefficient.

SELECTION ON THE BASIS OF CORRELATION

There are numerous methods of measuring the extent to which two series of numbers vary with each other — the

degree of the tendency for large numbers in one series to go with large numbers in the other series, and the tendency for small numbers in the first series to go with small numbers in the second. This is known as *concomitant variation*.¹ Among the more exact of these methods are those which result in what are known as *correlation coefficients*. Of the various correlation coefficients, the product-moment coefficient devised by Karl Pearson is by far the best. In fact, no other coefficient is adequate for use in the organization of aptitude batteries. For this reason no other coefficient will be presented in the present work.

There are numerous methods by which the Pearson or product-moment correlation coefficient may be computed.² For the most part, each method has special advantages which make it desirable for use in certain situations. For aptitude work the procedure about to be described is perhaps the best. It is easy to learn. It avoids the confusing combinations of plus and minus signs characteristic of some methods. No errors are introduced into the results by the grouping of data. Since the number of subjects in aptitude projects usually does not run above 150 and since test scores rarely need exceed two-place numbers, the numerical totals encountered hardly ever become inconveniently large. Lastly, the procedure provides a means of checking the accuracy of practically all the computations leading up to the correlation coefficients. This last is of the greatest importance, because an error of any considerable size in the computation at this point of an aptitude project would seriously impair the value of the entire investigation.

¹ See pages 10 ff.

² Fifty different formulæ for computing the product-moment coefficient of correlation have been brought together by Symonds (84).

METHOD OF COMPUTING CORRELATION COEFFICIENTS ILLUSTRATED

The basic principles of the method may be understood easily from the examination of one or two examples based on miniature sets of data. The first example will be the computation of the extent of the correlation between columns *A* and *B* of Table 71. The first thing to do is to square each item in column *A* and enter the results in column *A*² as shown.

TABLE 71

SHOWING THE PRELIMINARY COMPUTATIONS FOR A POSITIVE CORRELATION
COEFFICIENT, MINIATURE SET OF DATA

SUBJECT NO.	ORIGINAL DATA		PRELIMINARY COMPUTATIONS		
	<i>A</i>	<i>B</i>	<i>A</i> ²	<i>B</i> ²	<i>A</i> × <i>B</i>
I	13	9	169	81	117
II	2	8	4	64	16
III	4	5	16	25	20
IV	6	7	36	49	42
V	10	11	100	121	110
Sums . . .	35	40	325	340	305
Means . . .	7	8	65	68	61

The items of column *B* are squared in the same manner, the squares appearing in column *B*². Then each item in *B* is multiplied by the corresponding item in *A*, the results appearing in column *A* × *B*. Then all five columns are added and averaged. It is customary to call these averages *means*.

The correlation coefficient may now be computed by the following formula :

$$r = \frac{M_{A \times B} - M_A \times M_B}{\sqrt{M_{A^2} - (M_A)^2} \sqrt{M_{B^2} - (M_B)^2}} \quad . \quad . \quad (17)$$

In this formula ¹ the letter M stands for *mean* and the subscripts, or small letters placed at the right of each M and a little below, show which mean is meant. Thus M_A is the mean of column A , or 7; M_{A^2} is the mean of column A^2 , or 65; $M_{A \times B}$ is the mean of column $A \times B$, or 61. Substituting in the formula the various values from Table 71, we have:

$$r = \frac{61 - 7 \times 8}{\sqrt{65 - (7)^2} \sqrt{68 - (8)^2}}$$

Solving by easy stages:

$$r = \frac{61 - 56}{\sqrt{65 - 49} \sqrt{68 - 64}}$$

$$r = \frac{5}{\sqrt{16} \sqrt{4}}$$

$$r = \frac{5}{4 \times 2}$$

$$r = \frac{5}{8}$$

$$r = +.625$$

A second example (in this case illustrating a negative coefficient) may serve still further to fix the use of the formula. The data to be correlated are presented in columns A and B of Table 72. The two columns are first squared and the cross multiplications are carried out and entered in the appropriate columns, after which the sums and means are found just as described in the preceding example. Substituting in the formula, we have:

$$r = \frac{15 - 3 \times 7}{\sqrt{13 - 3^2} \sqrt{65 - 7^2}}$$

¹ This is a modification of a formula devised by J. A. Harris (28) and later discovered independently by L. L. Thurstone and still later by Leonard Ayres.

TABLE 72

SHOWING THE PRELIMINARY COMPUTATIONS FOR A NEGATIVE CORRELATION COEFFICIENT, MINIATURE SET OF DATA

SUBJECT NO.	ORIGINAL DATA		PRELIMINARY COMPUTATIONS		
	A	B	A ²	B ²	A × B
I	0	14	0	196	0
II	2	8	4	64	16
III	3	5	9	25	15
IV	4	2	16	4	8
V	6	6	36	36	36
Sums	15	35	65	325	75
Means . . .	3	7	13	65	15

Solving as before:

$$r = \frac{15 - 21}{\sqrt{13} - 9 \sqrt{65} - 49}$$

$$r = \frac{-6}{\sqrt{4} \sqrt{16}}$$

$$r = \frac{-6}{2 \times 4}$$

$$r = -\frac{6}{8}$$

$$r = -.75$$

It will be observed that in the numerator the multiplication is carried out before the subtraction. In case the product from this multiplication is larger than the $M_{A \times B}$ (as in the present example), the sign of the coefficient is minus.

It is important also to remember that the radicals in the denominator represent the standard deviations (S.D.'s or σ 's) of the respective columns of data. Thus

$$\sqrt{M_{A^2} - (M_A)^2} = \text{S.D.}_A \text{ or } \sigma_A \quad . \quad . \quad . \quad (18)$$

and $\sqrt{M_{B^2} - (M_B)^2} = \text{S.D.}_B \text{ or } \sigma_B$

These convenient symbols will be used later rather frequently, instead of the somewhat clumsy radicals.

SYSTEMATIC METHODS OF COMPUTATION NECESSARY

In aptitude work all possible correlations among the different variables must be computed. When the number of variables goes beyond three or four, the number of r 's required increases very rapidly, as may be seen by a glance at Table 73.

TABLE 73

SHOWING THE NUMBER OF CORRELATIONS REQUIRED BY VARYING
NUMBERS OF VARIABLES

NO. OF VARIABLES	NO. OF CORRELATIONS REQUIRED	TOTAL NUMBER COLUMNS OF SQUARES AND PRODUCTS TO BE COMPUTED
4	6	10
5	10	15
6	15	21
7	21	28
8	28	36
9	36	45
10	45	55
11	55	66
12	66	78
13	78	91
14	91	105
15	105	120
16	120	136

It therefore becomes both necessary and possible to economize the labor involved and at the same time practically to eliminate the liability to erroneous final results, by carrying out the work systematically. While the economies involved do not appear so strikingly as they would with a larger project, the procedure may best be explained by a miniature problem involving only four variables.

THE PRELIMINARY CORRELATION WORK SHEET

The first step in the procedure is to draw up on a sheet of paper of suitable size, preferably a good quality of white cardboard, the blank form of a preliminary correlation work sheet such as is shown in Table 74. On this work sheet are first entered the scores of the experimental data. The column of criterion scores is headed X_0 , and the columns of test scores are headed by X 's with subscripts from 1 on, consecutively. *The row of scores for each subject is then added, and the sums thus obtained are entered in the check column headed "ck."* The items in this column are then squared and the squares entered in the adjoining column headed " ck^2 ." Next the squares and the products of all the columns of experimental data are found and entered in the appropriate columns of the work sheet. These columns are headed by the combination of subscripts of the columns making them up. For example, 0×0 indicates column X_0 multiplied by itself, and 2×3 indicates X_2 multiplied by X_3 . Lastly, all the columns of figures are added and then averaged, the results being recorded at the bottom of the sheet in the rows labeled "Sums" and "Means" respectively. We now have things in shape, if we wished, to compute the various r 's by the formula as in the two examples worked out above. It will be much better, however, to do this systematically on special work sheets, as will be shown presently. But before doing this we must consider a valuable aid in the carrying out of the computations of the table just considered.

USE OF THE MULTIPLICATION TABLE

The finding of the squares and products constitutes decidedly the major part of the labor in aptitude computations. To realize this it is necessary only to recall that in a project involving ten variables there are no less than 55 such columns.

TABLE 74

PRELIMINARY CORRELATION WORK SHEET FOR THE SYSTEMATIC COMPUTATION OF ALL THE CORRELATION COEFFICIENTS REQUIRED IN AN APTITUDE PROJECT

Subject No.	EXPERIMENTAL DATA				CHECK COLUMNS		PRELIMINARY COMPUTATIONS FOR r										
	Criterion Score	Test Scores			Sums	Sums Squared	Squares				Products						
		X_0	X_1	X_2			X_3	0×0	1×1	2×2	3×3	0×1	0×2	0×3	1×2	1×3	2×3
1	13	9	0	14	36	1296	81	0	196	117	0	182	0	126	0		
2	2	8	2	8	20	400	64	4	64	16	4	16	16	64	16		
3	4	5	3	5	17	289	25	9	25	20	12	20	15	25	15		
4	6	7	4	2	19	361	39	16	4	42	24	12	28	14	8		
5	10	11	6	6	33	1089	121	36	36	110	60	60	66	66	36		
Sums	35	40	15	35	125	3435	340	65	325	305	100	290	125	295	75		
Means	7	8	3	7	25	687	68	13	65	61	20	58	25	59	15		

Checks: (A) $125 = 35 + 40 + 15 + 35$

(B) $25 = 7 + 8 + 3 + 7$

(C) $3435 = 325 + 340 + 65 + 325 + 2(305 + 100 + 290 + 125 + 295 + 75)$

(D) $687 = 65 + 68 + 13 + 65 + 2(61 + 20 + 58 + 25 + 59 + 15)$

If there are 100 subjects in the trial group, this means over five thousand multiplications. The time, labor, and liability of error involved in this procedure can be reduced enormously by the use of multiplication tables. If the data contain many three-place numbers, Crelle's *Rechentafeln* may be used. This is a multiplication table from which may be read off directly the products of all numbers up to 1000×1000 . Numbers larger than three places probably are never justified in aptitude work. Indeed, in most projects a little care in planning will keep both the criterion score and the test scores down to two-place numbers. This is not only very desirable for purposes of the computations involved, but is really required in the use of the resulting test battery, since large test scores are clumsy to manipulate in practical testing work. In such cases the multiplication table given in Appendix IV (pages 502-521) is far more convenient than Crelle's, as the latter is rather large to handle and requires much more turning of leaves.

When using a multiplication table on a preliminary work sheet for the correlation coefficient, one should look up a whole row of products at a time and thus save much needless labor. For example, in Table 74 one would find the page of the table showing the 13's, from which can be read off the products of 0×0 , 0×1 , 0×2 , and 0×3 , without turning any leaves whatsoever. If the project involved ten variables the economy would be over twice as great, since ten products could be read off without delay, instead of four.

THE STANDARD DEVIATION WORK SHEET

We may now proceed to the systematic solution of the formula for securing the r 's. The first step is to construct a work sheet by means of which may be computed simultaneously all the standard deviations which make up the denominators of the formulæ. This is shown in Table 75.

TABLE 75

WORK SHEET FOR STANDARD DEVIATIONS AND SQUARE MOMENTS

VARIABLE NUMBER	MEANS, M_A	MEANS SQUARED, $(M_A)^2$	MEANS OF THE SQUARES, M_A^2	SQUARE MOMENTS, $M_A^2 - (M_A)^2$	S.D. OR σ , $\sqrt{M_A^2 - (M_A)^2}$
0	7	49	65	16	4
1	8	64	68	4	2
2	3	9	13	4	2
3	7	49	65	16	4
Check Sums	25	171	211	40	12

Check: $(E) 40 = 211 - 171$

The items in the columns M_A and M_{A^2} are simply transferred from Table 74. The items in column M_A are squared and the squares entered in column $(M_A)^2$. Next, each one of these squares is subtracted from the corresponding item in the column M_A^2 , the difference being entered in the column of square moments. *The value M_A^2 is always larger than $(M_A)^2$ unless an error has been made.* Next the square root is taken of each of the square moments, the roots being entered in the column of S.D.'s or σ 's. Lastly, the columns of the work sheet are added and the sums entered at the foot of the several columns for purposes of checking.

THE USE OF TABLES FOR FINDING POWERS AND ROOTS

It is usually uneconomical to extract roots by hand. For all numbers under 1000 the square roots may be found directly from the table in Appendix III (pages 495-501) of the present volume. For larger numbers Barlow's tables should be used. Barlow gives the squares, cubes, square roots, cube roots, and reciprocals for all numbers up to 10,000. In the case of very large numbers, the square roots may be computed economically on a good calculating machine like the Monroe.

When looking up in the tables the square roots of numbers containing decimals, the beginner should be careful to remember that the number looked up in the table always must have an *even* number of decimal places. For example, if it were desired to find the square root of 9.6, it will be necessary to look up the square root of 960, which is 30.984, by the table. But since two decimal places in the power always correspond to one place in the root, we must move the point one place to the left, making the root of our number 3.0984. If 96 had been looked up in the table, the root would have been absolutely meaningless; similarly, if the number were 9.657. If a cipher be added, the number becomes 9.6570. This gives an even number of decimal places but makes the number too large even to be found in Barlow. We accordingly drop off the last decimal place, looking up in the table the number 966, since the 7 which is dropped is more than a half. In general, the dropping of decimals on powers produces very little effect on roots.

THE FINAL CORRELATION WORK SHEET

The form of the final work sheet for the correlation coefficients is shown in Table 76. The pairs of variable combinations are arranged systematically according to number in the first column of the sheet. The corresponding standard deviations from the S.D. work sheet are entered in columns headed respectively S.D._A and S.D._B. The corresponding means are entered in the columns headed respectively M_A and M_B. The values in column M_{A×B} are taken from the first or preliminary work sheet (Table 74).

Computation now begins. The pairs of S.D.'s are first multiplied one after the other, the products being entered in column S.D._A × S.D._B. This product represents the denominator of the formula. In a similar manner the pairs of means are multiplied, one after the other, the products being

TABLE 76
FINAL WORK SHEET FOR CORRELATION COEFFICIENTS

VARIABLE COMBINATIONS	DENOMINATOR			NUMERATOR				r $\frac{M_A \times B - M_A \times M_B}{S.D.A \times S.D.B}$
	S.D.A	S.D.B	S.D.A $\times$ S.D.B	M_B	$M_A \times M_B$	$M_A \times B$	Product Moment $M_A \times B - M_A \times M_B$	
0 and 1	4	2	8	8	56	61	+ 5	+ .625
0 and 2	4	2	8	3	21	20	- 1	- .125
0 and 3	4	4	16	7	49	58	+ 9	+ .5625
1 and 2	2	2	4	3	24	25	+ 1	+ .25
1 and 3	2	4	8	7	56	59	+ 3	+ .375
2 and 3	2	4	8	7	21	15	- 6	- .75
Check Sums			52		227	238	+ 11	

Checks: (F) $12^2 - 40 = 2 \times 52$
(G) $25^2 - 171 = 2 \times 227$
(H) $238 - 227 = 11$

entered in the column headed $M_A \times M_B$. Next the entries in column $M_A \times M_B$ are subtracted from the corresponding entries in column $M_{A \times B}$ and the difference recorded with appropriate sign in the numerator or product-moment column, headed $M_{A \times B} - M_A \times M_B$. If $M_{A \times B}$ is larger, the sign is plus; if $M_A \times M_B$ is larger, the sign is minus. The entry in the moment (or numerator) column is then divided by the corresponding entry in the denominator ($S.D._A \times S.D._B$) column. The resulting quotient is the correlation coefficient sought. This is recorded with appropriate sign in the final column of the work sheet. Lastly, all the columns of results computed on the present work sheet, except the r 's, are added. The sums are recorded beneath the respective columns.

HOW TO GUARANTEE ABSOLUTE ACCURACY OF THE CORRELATION COEFFICIENTS

We must now consider the extremely important practical matter of making sure that the various steps in the computations of our correlation coefficients, as outlined above, are absolutely without error. Persons who have not had considerable experience in computing large numbers of correlations such as are required in aptitude work usually do not realize the amount of inaccuracy usually present in such calculations when they are carried out by ordinary methods. Yet for the final choice of the tests of the battery to be made on a sound basis and for the weights assigned to the tests thus chosen to be usefully accurate, every individual coefficient must be correct. A single erroneous coefficient may ruin the entire project. The next few pages will show how absolute accuracy of correlation computations may be assured.

The basis of our method of securing strict accuracy in the computation of the correlation coefficients has already been

given in the eight checks shown at the foot of Tables 74, 75, and 76. For purposes of convenience these checks have been designated *A, B, C, D, E, F, G, and H* respectively. A ninth check, somewhat different in principle from the others and so not given a letter, completes the checking procedure. In some ways the simplest method of learning the use of these checks is to study them in the concrete as they appear beneath the various tables and then apply them to a new problem by simple analogy. The reader is accordingly advised to follow this method, at least in part. As a supplement to the above method, however, the following explanation is given :

Check A

Check *A* may be stated as follows : *The sum of the sums of the original data columns must equal the sum of the ck column.* Thus in Table 74 we have :

$$125 = 32 + 40 + 15 + 35$$

$$125 = 125$$

Since check *A* is not influenced by dropped decimals, if any disagreement whatever is found it indicates an error which must be located and corrected. Obviously such an error must be either in the addition of the rows or of the columns. If an adding machine has been used, the adding-machine slips should be checked with the original data to find where a wrong key has been pressed. *These adding-machine slips should show an asterisk at the top of each series added*, to make it certain that no numbers remaining in the machine from some previous work inadvertently have been included in some of the totals.

Check B

In case check *A* comes out correctly, check *B* may be applied. It may be stated as follows: *The sum of the means of the columns of original data (except for dropped decimals) must equal the mean of the check column.* Thus in Table 74 we have:

$$25 = 7 + 8 + 3 + 7$$

$$25 = 25$$

In case this check fails (except for minor discrepancies due to dropped decimals), the means must be recomputed and corrected until the check is perfect.

Check C

Check *C*, easily the most important of all, may be stated thus: *The total of the square sums plus twice the total of the product sums must equal the sum of the ck^2 column.* Thus in Table 74 we find:

$$3435 = 325 + 340 + 65 + 325$$

$$+ 2(305 + 100 + 290 + 125 + 295 + 75)$$

$$3435 = 1055 + 2 \times 1190$$

$$3235 = 1055 + 2380$$

$$3435 = 3435$$

Since check *C* is not influenced by dropped decimals, it must be exact or an error has been made. The error may lie either in the additions or in the multiplications. It is usually best to check the additions first. In case an adding machine has been used, the additions may be checked merely by comparing the adding-machine slips with the corresponding columns of the work sheet.

If the additions are found correct, the error must be sought in the multiplications — either in the squares or the products. This may be done in either of two ways. Probably the simplest method is to apply to the various *rows* of squares

and products in the body of the table the method already observed in check *C* as applied to the sums. For example, in the case of Table 74, first row of multiplications :

$$\begin{aligned} 1296 &= 169 + 81 + 0 + 196 \\ &\quad + 2(117 + 0 + 182 + 0 + 126 + 0) \\ 1296 &= 446 + 2(425) \\ 1296 &= 1296 \end{aligned}$$

This procedure may be carried out very rapidly by means of an adding machine, each square being added once and each product twice. When a row is encountered which fails to check, the multiplications of that row must be performed anew until the error is found and all corrections have been made. After this, check *C* should be applied once more to make sure that no other errors exist.

The second or Ezekiel method of locating errors in multiplication¹ depends on the principle that the sum of the products of any column of original data multiplied by the check column equals the sum of the sums of the products of that column multiplied by each of the other columns together with the sum of its own squares. For example, the sum of the products of column 0 multiplied by the check column must be equal to the sum of the sums of the products of 0×0 , 0×1 , 0×2 , and 0×3 . Similarly, the sum of column 1 multiplied by the check column must be equal to the sum of the sums of the products of 1×0 , 1×1 , 1×2 , and 1×3 , and so on for the other combinations. In case one of these series fails to check, then the multiplications involved must be repeated until the error is found and the check is perfect.

¹ The attention of the writer was called to this method in 1925 by Professor P. E. McNall of the Wisconsin College of Agriculture. It is said to have been originated by Dr. Mordecai Ezekiel, though he seems not to have published any account of it.

Check D

Check *D* may be stated as follows: *The total of the square means plus twice the total of the product means must equal (except for dropped decimals) the mean of the ck^2 column.* Thus in Table 74 we find:

$$\begin{aligned} 687 &= 65 + 68 + 13 + 65 \\ &\quad + 2(61 + 20 + 58 + 25 + 59 + 15) \\ 687 &= 211 + 2(238) \\ 687 &= 687 \end{aligned}$$

The means of all columns should be carried out to three decimal places, so that this check should be pretty nearly exact. In case of a discrepancy larger than would result from dropped decimals, the means should be recomputed until the check is satisfactory.

Check E

Check *E* is as follows: *The sum of column $(M_A)^2$ subtracted from the sum of column M_A^2 must equal the sum of the column of square moments.* Thus in Table 75 we have:

$$\begin{aligned} 40 &= 211 - 171 \\ 40 &= 40 \end{aligned}$$

If this check fails, the difficulty must be sought in the subtractions by which the several entries in the column of square moments were obtained.

Check F

Check *F* is as follows: *The square of the sum of the σ 's minus the sum of the square moments equals twice the sum of the σ products.* Assembling the relevant data from Tables 75 and 76, we have:

$$\begin{aligned} 12^2 - 40 &= 2 \times 52 \\ 144 - 40 &= 104 \\ 104 &= 104 \end{aligned}$$

This check tests the accuracy of several different operations.

In case an error is disclosed, it may be in any one of the following steps:

1. The extraction of the roots of the square moments which yielded the σ 's.
2. The addition of the σ 's.
3. The copying in of the σ values into the final work sheet.
4. Multiplication of the pairs of σ 's.
5. The addition of the column of σ products.

Check G

Check G runs as follows: *The square of the sum of the means minus the sum of the squares of the means equals twice the sum of the mean products.* Taking illustrative material from Tables 75 and 76, we have:

$$25^2 - 171 = 2 \times 227$$

$$625 - 171 = 454$$

$$454 = 454$$

Just as with check F, this check also tests the accuracy of several different operations. Accordingly, in case the check reveals an error, it may be found in any one of the following steps:

1. The squaring of the means.
2. The additions of the mean-squares.
3. The transfer of the means to the final work sheet.
4. The multiplication of the pairs of means.
5. The addition of the mean-products.

Check H

Check H may be stated thus: *The sum of the column of mean-products ($M_A \times M_B$) subtracted from the sum of column ($M_{A \times B}$) equals the algebraic sum of the product moments.* Thus from Table 76 we have:

$$238 - 227 = + 11$$

$$+ 11 = + 11$$

Of the above values, the 238 has already been tested in check *B* and the 227 in check *E*. Accordingly, if an error has been made it can be only in the subtraction resulting in the product moments. This, of course, may be either in the absolute values of the numbers or in the signs.

There now remains only to check the final operation which produced the correlation coefficients. No special check analogous to the eight already described for other parts of the procedure is available here. Accordingly each r must be checked separately by repeating the operation in reverse. A good way to do this is to multiply the r by the corresponding σ -product. This should yield the product moment, except for decimals which may have been dropped :

$$+.625 \times 8 = +5.000$$

Every individual r should be checked in this way in order that absolute accuracy may be assured.

CORRELATION TECHNIQUE OF FREE-HAND-DRAWING PROJECT

With the mechanical aspects of the correlation technique disposed of, we may now proceed to an exposition of its use in the evaluation of the test units of a preliminary battery. This can best be done in connection with the consideration of a concrete example. We have chosen for this purpose the free-hand-drawing project, the preceding steps of which have already been described in some detail in previous chapters. The final criterion scores appear in Table 70. The first thing is to drop off excess decimals in certain columns of the original test scores, so that nothing beyond two-place numbers need be handled. Where the decimal dropped is 5 or above, the next retained digit is increased by one point. The experimental data ready for computation, together with the two check columns, are assembled in Table 77. The scores of the various tests may be identified by the subscripts of the X 's

TABLE 77

SHOWING THE EXPERIMENTAL DATA OF THE FREE-HAND-DRAWING APTITUDE PROJECT, READY FOR COMPUTATION. SUFFICIENT DECIMALS HAVE BEEN DROPPED TO REDUCE PRACTICALLY ALL NUMBERS TO TWO PLACES OR LESS.

SUBJECT NUM- BER	EXPERIMENTAL DATA										CHECK COL- UMN SUMS ¹	CHECK COL- UMN SQUARES
	Criter- ion Score	Test Scores										
		X ₀	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈		
1	93	11	6	0	6	10	8	1.6	1.6	.23	189	35721
2	72	23	7	0	22	18	3	3.4	5.8	.62	299	89401
3	90	23	10	0	14	7	14	4.6	2.6	.33	263	69169
4	79	22	16	1	44	16	9	3.7	2.6	.44	294	86436
5	79	20	12	4	8	31	3	1.5	2.8	.79	279	77841
6	89	21	2	0	18	29	8	2.1	3.6	.25	249	62001
7	90	20	7	0	6	12	4	4.2	2.6	.31	238	56644
8	91	22	13	0	54	19	5	6.1	3.1	.56	352	123904
9	82	13	9	0	12	21	6	5.0	4.8	.73	314	98596
10	80	15	12	1	6	8	6	3.1	2.9	.27	215	46225
11	80	14	14	0	24	18	4	2.9	3.6	.40	259	67081
12	80	22	13	0	30	14	8	3.6	4.3	.42	288	82944
13	89	19	10	1	10	18	4	3.4	3.1	.25	241	58081
14	83	20	10	0	4	28	11	1.3	4.0	1.12	321	103041
15	76	17	7	0	12	29	3	3.7	2.0	.62	263	69169
16	87	21	9	0	2	15	9	5.7	3.1	.40	271	73441
17	91	19	13	1	6	20	17	3.3	5.4	.10	264	69696
18	71	7	8	0	14	33	3	2.1	3.0	.40	227	51529
19	65	18	10	0	12	17	3	2.4	7.0	.48	267	71289
20	81	12	13	0	26	30	5	1.8	5.7	.37	279	77841
21	74	21	4	1	15	27	10	2.8	3.4	.31	245	60025
22	68	18	4	0	28	13	3	2.9	3.7	.40	240	57600
23	85	13	13	0	15	24	16	2.1	2.3	.29	239	57121
24	85	21	8	0	36	17	1	1.4	2.6	.08	216	46656
25	82	21	10	1	10	18	7	2.8	5.3	.35	265	70225
26	69	15	1	1	50	30	4	4.1	5.9	.35	305	93025
27	75	9	6	1	24	22	16	3.6	3.9	.67	295	87025
28	83	20	18	0	18	34	10	2.8	3.4	.52	297	88209
29	97	22	4	3	45	15	8	4.9	2.9	.25	297	88209
30	73	21	8	0	8	15	3	6.1	1.8	.54	261	68121
31	78	19	12	0	4	8	4	2.3	4.5	.40	233	54289
32	82	23	16	1	31	16	14	4.1	3.8	.37	299	89401
33	87	17	16	0	30	13	11	4.2	3.2	.17	265	70225
34	89	22	5	0	5	24	7	3.1	2.7	.46	256	65536
35	80	21	18	0	10	21	13	3.2	1.7	.19	231	53361
36	80	12	4	1	7	13	6	2.8	3.7	.23	211	44521
37	71	16	6	1	8	18	5	3.3	1.2	.77	247	61009
38	73	14	6	0	27	26	0	3.2	2.7	.42	247	61009
39	77	15	0	3	12	13	4	3.6	5.3	.31	244	59536
40	73	21	7	0	19	33	6	5.1	5.1	.52	313	97969
41	87	12	13	0	5	14	20	6.2	2.9	.25	267	71289
42	79	21	10	0	9	27	6	2.3	2.2	.54	251	63001
43	85	22	17	0	8	15	12	3.4	1.6	.21	230	52900
44	80	22	9	1	6	49	6	4.7	4.0	.69	329	108241
45	78	20	9	1	9	24	8	4.1	6.3	.77	330	108900
46	77	18	9	0	5	29	9	4.1	3.4	.19	241	58081
47	80	20	9	0	6	13	6	6.1	5.0	.40	285	81225
48	88	23	7	1	24	15	9	.8	3.0	.35	240	57600
Sums	3883	878	450	24	804	979	357	165.6	171.1	20.09	12751	3444359
Means	80.9	18.3	9.375	.5	16.75	20.4	7.44	3.45	3.56	.42		

¹ In using the checking system all numbers are treated as whole numbers.

heading the columns, as explained on page 373. A preliminary work sheet 18" $\times$ 35" was drawn up on white cardboard according to the plan of Table 74. In this table were recorded the various squares and products of the preliminary computations, after which the various columns were averaged. The means thus obtained were then tested for accuracy of computation by checks *A*, *B*, *C*, and *D*. Following this, the various standard deviations and correlation coefficients were computed by means of their respective work sheets. These, together with the accompanying checks, are shown in Tables 78 and 79.

THE PROBLEM OF SCORING ERRORS IN TEST PERFORMANCE

With the technique and results of correlation fairly before us, it is now possible to consider in a satisfactory manner the troublesome subject of errors in test responses in relation to

TABLE 78

WORK SHEET FOR STANDARD DEVIATIONS AND SQUARE MOMENTS OF
FREE-HAND-DRAWING APTITUDE PROJECT

VARIABLE NUMBER	M_A	$(M_A)^2$	M_A^2	SQUARE MOMENTS, $M_A^2 - (M_A^2)$	S.D.'s, $\sqrt{M_A^2 - (M_A^2)}$
0	80.896	6544.163	6593.937	49.774	7.055
1	18.292	334.597	351.042	16.444	4.055
2	9.375	87.891	106.417	18.526	4.304
3	.5	.25	1.00	.75	.866
4	16.75	280.563	444.917	164.354	12.820
5	20.396	415.997	485.104	69.107	8.313
6	7.438	55.316	74.521	19.204	4.382
7	3.45	11.903	13.664	1.761	1.327
8	3.565	12.706	14.499	1.793	1.339
9	.4185	.1751	.2165	.0413	.203
Sums	161.0795	7743.560	8085.317	341.757	44.664

Check: (E) $341.757 = 8085.317 - 7743.560$
 $341.757 = 341.757$

TABLE 79

WORK SHEET FOR CORRELATION COEFFICIENTS OF THE FREE-HAND-
DRAWING APTITUDE PROJECT

VARIABLES NUMBER	DENOMINATOR			NUMERATOR							Cross Product Moments $M_A \times B - M_A \times M_B$	r
	σ_A	σ_B	$\sigma_A \times \sigma_B$	M_A	M_B	$M_A \times M_B$	$\Sigma A \times B$	$M_A \times B$				
0 and 1	7.055	4.055	28.608	80.896	18.292	1479.750	71343	1486.313	+ 6.563	+ .229		
0 " 2	7.055	4.304	30.365	80.896	9.375	758.4	36707	764.729	+ 6.329	+ .208		
0 " 3	7.055	.866	6.110	80.896	.5	40.448	1956	40.75	+ .302	+ .049		
0 " 4	7.055	12.820	90.445	80.896	16.75	1355.008	65050	1355.208	+ .2003	+ .002		
0 " 5	7.055	8.313	58.648	80.896	20.396	1649.945	78469	1634.771	- 15.184	- .259		
0 " 6	7.055	4.382	30.915	80.896	7.438	601.664	29495	614.479	+ 12.815	+ .415		
0 " 7	7.055	1.327	9.362	80.896	3.45	279.091	13426.9	279.727	+ .6358	+ .068		
0 " 8	7.055	1.339	9.447	80.896	3.565	288.362	13695.9	285.331	- 3.0306	- .321		
0 " 9	7.055	.203	1.432	80.896	.419	33.855	1601.37	33.362	- .493	- .344		
1 " 2	4.055	4.304	17.453	18.292	9.375	171.488	8373	174.438	+ 2.95	+ .169		
1 " 3	4.055	.866	3.512	18.292	.5	9.146	448	9.333	+ .1873	+ .053		
1 " 4	4.055	12.820	51.985	18.292	16.75	306.391	14957	311.604	+ 5.213	+ .100		
1 " 5	4.055	8.313	33.709	18.292	20.396	373.084	17841	371.688	- 1.396	- .041		
1 " 6	4.055	4.382	17.769	18.292	7.438	136.047	6533	136.104	+ .0574	+ .003		
1 " 7	4.055	1.327	5.381	18.292	3.45	63.107	3068.5	63.927	+ .8197	+ .152		
1 " 8	4.055	1.339	5.430	18.292	3.565	65.204	3124.5	65.094	- 1.099	- .020		
1 " 9	4.055	.203	.823	18.292	.419	7.655	369.56	7.699	+ .0439	+ .053		
2 " 3	4.304	.866	3.727	9.375	.5	4.688	183.	3.813	- .875	- .235		
2 " 4	4.304	12.820	55.177	9.375	16.75	157.031	7516	156.583	- .4479	- .008		
2 " 5	4.304	8.313	35.779	9.375	20.396	191.213	9068	188.917	- 2.296	- .064		
2 " 6	4.304	4.382	18.860	9.375	7.438	69.727	3711.	77.313	+ 7.5859	+ .402		
2 " 7	4.304	1.327	5.711	9.375	3.45	32.344	1561.6	32.533	+ .1895	+ .033		
2 " 8	4.304	1.339	5.763	9.375	3.565	33.418	1561.4	32.529	- .8889	- .154		
2 " 9	4.304	.203	.874	9.375	.419	3.923	185.89	3.873	- .0507	- .058		
3 " 4	.866	12.820	11.102	.5	16.75	8.375	453	9.438	+ 1.0625	+ .096		
3 " 5	.866	8.313	7.199	.5	20.396	10.198	502	10.458	+ .260	+ .036		
3 " 6	.866	4.382	3.795	.5	7.438	3.719	169	3.521	- 1.1979	- .052		
3 " 7	.866	1.327	1.149	.5	3.45	1.725	78.1	1.627	- .0979	- .085		
3 " 8	.866	1.339	1.160	.5	3.565	1.782	90.3	1.881	+ .0989	+ .085		
3 " 9	.866	.203	.176	.5	.419	.209	10.76	.224	+ .0149	+ .085		
4 " 5	12.820	8.313	106.573	16.75	20.396	341.633	16364	340.917	- .716	- .007		
4 " 6	12.82	4.382	56.177	16.75	7.438	124.578	5632	117.333	- 7.245	- .129		
4 " 7	12.82	1.327	17.012	16.75	3.45	57.788	2838.8	59.142	+ 1.354	+ .080		
4 " 8	12.82	1.339	17.166	16.75	3.565	59.707	2940.8	61.267	+ 1.560	+ .091		
4 " 9	12.82	.203	2.602	16.75	.419	7.010	322.08	6.71	- .2998	- .115		
5 " 6	8.313	4.382	36.428	20.396	7.438	151.695	7097	147.854	- 3.841	- .105		
5 " 7	8.313	1.327	11.031	20.396	3.45	70.366	3308.6	68.929	- 1.437	- .130		
5 " 8	8.313	1.339	11.131	20.396	3.565	72.704	3559.2	74.15	+ 1.447	+ .130		
5 " 9	8.313	.203	1.688	20.396	.419	8.536	441.54	9.199	+ .663	+ .393		
6 " 7	4.382	1.327	5.815	7.438	3.45	25.695	1272.7	26.515	+ .855	+ .147		
6 " 8	4.382	1.339	5.867	7.438	3.565	26.512	1238.7	25.806	- .705	- .120		
6 " 9	4.382	.203	.890	7.438	.419	3.113	141.28	2.943	- .169	- .190		
7 " 8	1.327	1.339	1.777	3.45	3.565	12.298	594.03	12.376	+ .078	+ .044		
7 " 9	1.327	.203	.269	3.45	.419	1.444	69.515	1.448	+ .004	+ .016		
8 " 9	1.339	.203	.272	3.565	.419	1.492	73.579	1.533	+ .041	+ .015		

the choice of the units to make up the final battery. Many tests, such as No. 2 (Directions test) of our preliminary battery for free-hand drawing, yield two scores — the number of items performed correctly and the number in which an error of some kind is made. These two scores are usually called the “rights” and the “wrongs,” respectively. Numerous arbitrary formulæ and procedures have been proposed for scoring errors in such a way as to combine them with the rights and thus secure a single comprehensive score for the test. In case of time-limit tests it is sometimes supposed that the subject has been sufficiently penalized by the mere loss of the time wasted on the incorrect item, so that the rights alone are counted. One of the most common methods is to deduct one correct response for every incorrect one. Sometimes two correct items are deducted for every error. A number of still more complicated methods have been suggested.

All such arbitrary procedures are unsound. It almost seems as if the idea behind them in some cases were one of retribution — that the subject had done wrong and must therefore be punished by having his score diminished. The beginner in aptitude work will do well to rid himself at once of any such notion. It is hardly necessary to point out that in the scoring of errors there is no question of fairness or justice to the subject. *The problem is simply the determination of what weight should be given to the errors so that when combined with the rights or with the other tests of a battery, the joint score thus secured will yield the maximum correlation with the criterion.* Moreover, if the errors have any significance at all, they can hardly be expected to have the same significance for all aptitudes. Just what the optimum weighting should be in any particular case can only be determined accurately by means of experiment, supplemented by appropriate statistical analysis.

HOW TO TELL WHETHER SCORED ERRORS WILL INCREASE
THE STRENGTH OF A TEST

As a matter of fact, errors in test performance are of much less importance than is often supposed. It is the experience of most workers in this field that the prognostic value of a test is rarely increased very much by combining the wrongs with the rights, over the yield from the rights alone. This weakness of test-error scores in prognostic batteries is probably due to the fact that in most tests the incorrect responses are so few in number that whatever complex of factors may be producing the errors is very inadequately sampled. Accordingly, before bothering much about the scoring of errors in any particular case, it is best to determine how much higher the test will correlate with the criterion if the errors should be scored in the best possible manner. Fortunately, this may be determined without the actual weighting of the errors. It is accomplished by means of Thurstone's formula (91):

$$R_{C(RW)} = \sqrt{\frac{r_{CR}^2 + r_{CW}^2 - 2 r_{CR} r_{CW} r_{RW}}{1 - r_{RW}^2}} \quad (19)$$

In this formula the subscripts *C*, *R*, and *W* refer to the *criterion*, the *rights*, and the *wrights*, respectively. $R_{C(RW)}$ is the correlation of the criterion with the rights and the wrongs when the latter are combined in the most effective way possible, from the point of view of estimating the criterion.

The use of this formula may be illustrated by applying it as a test to determine whether it will be worth while to score the errors on the directions test of the preliminary battery for free-hand drawing. The rights are listed as X_2 and the wrongs as X_3 . The three correlation coefficients involved in the formula are found in Table 79. Since nothing more than an approximation is desired, we shall take the coefficients merely to the nearest two-place numbers and thus

materially reduce the labor of computation. Substituting in the formula, we have :

$$R_{C(RW)} = \sqrt{\frac{.21^2 + .05^2 - 2 \times .21 \times .05 \times -.24}{1 - .24^2}}$$

The values of the two squares in the numerator may be looked up in the table of Appendix III (page 495) if they are large enough to make it worth while. The value of the entire expression in the denominator may be found in the table of Appendix II.

$$\begin{aligned} R_{C(RW)} &= \sqrt{\frac{.0441 + .0025 + .00504}{.94}} \\ &= \sqrt{.0548} \end{aligned}$$

Looking up the root of .0548 in Appendix III,

$$R_{C(RW)} = .234$$

Since we began the computation with the assumption that r_{CW} was .21, it is seen that at the very best the wrongs should add only about two points to the correlation given by the rights alone. It is perfectly evident, then, that such a small gain would not be worth the time and labor necessary to secure it. Indeed, it probably would not have been worth while unless the correlation had been increased by five or six points at the least. It is true, of course, that the contribution required as the condition for the scoring of errors on a test already in a battery and administered, naturally is much less than that required for the inclusion of a separate test.

HOW TO SCORE ERRORS IF RETAINED FOR USE IN THE FINAL BATTERY

If the errors of a test turn out upon investigation to be worth scoring, either of two methods may be followed. The first is to determine the weight to be given the errors such

that, when combined with the rights, the two taken by themselves will yield the maximum correlation with the criterion. This may be accomplished by Thurstone's formula (91) :

$$\text{Weight of wrongs} = \frac{\sigma_R(r_{CR} \times r_{RW} - r_{CW})}{\sigma_W(r_{CW} \times r_{RW} - r_{CR})} \quad (20)$$

Here the subscripts mean the same as in the formula given above. When the weight of the wrongs is found, then the number of errors made by each subject is multiplied by it. If the value of the weight turns out to be negative, the product is subtracted from the score of the rights. But if the weight turns out to have a positive sign, then the product is *added* to the score of rights. The formula thus not only yields the best weight to be given the wrongs, but indicates whether they are to be added to, or subtracted from, the rights.

As an illustration of the use of this convenient formula, we shall apply it to the directions test just considered. The necessary correlation coefficients are given in Table 79 and the σ 's in Table 78. Substituting, we have

$$\text{Weight of wrongs} = \frac{4.3(.21 \times -.24 - .05)}{.87(.05 \times -.24 - .21)}$$

Solving by easy stages, being careful of signs :

$$\begin{aligned} \text{Weight of wrongs} &= \frac{4.3(-.0504 - .05)}{.87(-.012 - .21)} \\ &= \frac{4.3(-.1004)}{.87(-.222)} \\ &= \frac{-.432}{-.193} \\ &= + 2.2 \end{aligned}$$

This means that in the present case the errors should be *added* to the rights instead of subtracted, if the best results would be obtained from their use.

The actual employment of the weight of $+2.2$ as found by the above formula may be illustrated by the following example: Suppose a subject on the directions test gets 15 items right and 3 wrong. We simply multiply 3 by 2.2, which equals 6.6. This is *added* to the 16, which yields 21.6 as the combined score. The method applied systematically to the relevant data of the free-hand-drawing project is shown in Table 80. Columns X_2 and X_3 have been transferred from Table 77. By actual computation the column of combined scores where 2.2 times the wrongs was added correlates $+.233$ with the criterion, which agrees exactly with the value derived from Formula 19, except for dropped decimals. By way of contrast there is given in the last column of Table 80 the result of combining the rights with the wrongs by the method usually employed; i.e., by subtracting the wrongs from the rights. By actual computation this column of combined scores correlates with the criterion only to the extent of $+.186$, which is actually less than the correlation yielded by the rights alone. This serves to illustrate the fact that the weighting of errors by guess, instead of strengthening a test, may actually injure it.

GENERAL CORRELATIONAL CHARACTERISTICS DESIRABLE OF
TESTS FOR A BATTERY

With a complete set of correlations computed and the problem of scoring errors disposed of, the choice of the tests for the final battery may be made. The choice may be performed rather simply by a mere inspection of the coefficients, or it may become rather elaborate and employ the technique of partial correlation. After one has had some experience in organizing test batteries, he usually develops considerable skill in the interpretation of such masses of correlation coefficients so that partial correlation rarely needs to be resorted to. A number of principles, however, them-

TABLE 80

SHOWING THE METHOD OF COMBINING TEST ERRORS WITH
RIGHTS WHEN THE WEIGHT OF THE WRONGS IS + 2.2

SUBJECT No.	RIGHTS (X_2)	WRONGS (X_1)	WRONGS AFTER WEIGHTING	RIGHTS PLUS WEIGHTED WRONGS	RIGHTS MINUS WRONGS
1	6	0	+ .0	6.0	6.0
2	7	0	+ .0	7.0	7.0
3	10	0	+ .0	10.0	10.0
4	16	1	+ 2.2	18.2	15.0
5	12	4	+ 8.8	20.8	8.0
6	2	0	+ .0	2.0	2.0
7	7	0	+ .0	7.0	7.0
8	13	0	+ .0	13.0	13.0
9	9	0	+ .0	9.0	9.0
10	12	1	+ 2.2	14.2	11.0
11	14	0	+ .0	14.0	14.0
12	13	0	+ .0	13.0	13.0
13	10	1	+ 2.2	12.2	9.0
14	10	0	+ .0	10.0	10.0
15	7	0	+ .0	7.0	7.0
16	9	0	+ .0	9.0	9.0
17	13	1	+ 2.2	15.2	12.0
18	8	0	+ .0	8.0	8.0
19	10	0	+ .0	10.0	10.0
20	13	0	+ .0	13.0	13.0
21	4	1	+ 2.2	6.2	3.0
22	4	0	+ .0	4.0	4.0
23	13	0	+ .0	13.0	13.0
24	8	0	+ .0	8.0	8.0
25	10	1	+ 2.2	12.2	9.0
26	1	1	+ 2.2	3.2	0.0
27	6	1	+ 2.2	8.2	5.0
28	18	0	+ .0	17.2	18.0
29	4	3	+ 6.6	10.6	1.0
30	8	0	+ .0	8.0	8.0
31	12	0	+ .0	12.0	12.0
32	16	1	+ 2.2	18.2	15.0
33	16	0	+ .0	16.0	16.0
34	5	0	+ .0	5.0	5.0
35	18	0	+ .0	18.0	18.0
36	4	1	+ 2.2	6.2	3.0
37	6	1	+ 2.2	8.2	5.0
38	6	0	+ .0	6.0	6.0
39	0	3	+ 6.6	6.6	- 3.0
40	7	0	+ .0	7.0	7.0
41	13	0	+ .0	13.0	13.0
42	10	0	+ .0	10.0	10.0
43	17	0	+ .0	17.0	17.0
44	9	1	+ 2.2	11.2	8.0
45	9	1	+ 2.2	11.2	8.0
46	9	0	+ .0	9.0	9.0
47	9	0	+ .0	9.0	9.0
48	7	1	+ 2.2	9.2	6.0

selves derived from partial and multiple correlation, will be useful to the beginner.

If a single test were to be employed for purposes of forecasting, the situation would be very simple indeed. The only consideration apart from the expense incidental to the use of the test would be that it should correlate as highly as possible with the criterion. This principle is of such great and obvious significance that some of the early psychologists working with tests made the mistake of employing it exclusively, even when choosing a number of tests to make up a battery.

Where several tests are to be combined into a battery there must be considered, not only the correlation of each test with the criterion, but also the correlation of each test with the other tests. This fact is embodied in the useful though inexact maxim which states that *the correlation between the tests and the criterion should be as large as possible, whereas the correlations among the tests should be as small as possible*. The theoretical basis of this maxim has been considered in an earlier chapter (pages 254 ff. and 271 ff.). It remains here to point out certain details and exceptions not easily incorporated into a general statement. These details and exceptions relate largely to the signs of the correlation coefficients.

SIGNIFICANCE OF CORRELATION SIGNS IN CHOICE OF TESTS

In this connection it should be noted that if all variables in aptitude work were so constituted that a large score would represent a good performance, nearly all coefficients in aptitude work would be positive. There would rarely be negative coefficients of any considerable size. Accordingly, the signs of the coefficients in aptitude work are usually significant of nothing but the mode of scoring the variables involved. Thus, if a *criterion* has been so scored that the larger the score the better the performance, then a test given by the *time-limit method* will ordinarily yield a positive correlation with

the criterion because, by this method of scoring, the larger the test score, also, the better the performance.

But if the test has been given by the *work-limit method*, the coefficient with the criterion assumed above will ordinarily be negative. This is because the *weaker* the subject, the longer the time required to perform the task and hence the larger the test score. Accordingly, in this case the larger the test score, the *worse* the performance; hence the negative correlation with the criterion.

FOUR TYPE CORRELATIONAL COMBINATIONS CONSIDERED AS TO BATTERY DESIRABILITY

This brings us to the consideration of specific cases of what is desirable in the correlation between tests for different types of criterion correlation combinations. In the following paragraphs it will be understood that the expressions "high" or "low" when referring to correlations express the joint significance of the sign as well as the size of the coefficient, whereas "large" and "small" indicate only the absolute size. Thus of the two coefficients: $+.10$, $-.15$, the second is *lower* but the first is *smaller*.

CASE I. If the correlations between each of two tests and the criterion are both positive and large, the tests ordinarily will correlate positively with each other. In this case *it is desirable that the correlation between the tests shall be as low as possible*. This principle is illustrated by the series of yields resulting from combining two tests with fixed and rather high criterion correlations but with varying correlations between themselves. Such a series is shown in Table 81. The arrow indicates the direction of desirability. From the column of yields it is evident that the most desirable correlations between tests in this case are negative, and if negative, the larger the better. Unfortunately, human nature is so constituted that with the criterion correlations as assumed,

TABLE 81

SHOWING FOR A PAIR OF TESTS HAVING HIGH POSITIVE CRITERION CORRELATIONS, THE CORRELATION YIELDS RESULTING FROM COMBINING THEM UNDER VARYING ASSUMPTIONS AS TO THE CORRELATION WITH EACH OTHER

CRITERION CORRELATIONS, CASE I	VARYING TEST CORRELATIONS (r_{12})	DIRECTION OF DESIRABILITY OF TEST CORRELATIONS	CORRESPONDING YIELD OF COMBINED TESTS (R)
$r_{01} = +.50$ $r_{02} = +.40$	+.70	↓	.501
	+.40		.546
	+.20		.588
	.00		.640
	-.05		.658
	-.10		.673

reliable correlations between tests extending any distance below zero are rarely encountered.

CASE II. If the correlations between two tests and a criterion are both negative and large, they will ordinarily correlate positively with each other. Exactly as in Case I, in Case II it is also desirable that the correlation between the tests shall be as low as possible. The situation with respect to yields is shown in Table 82, the arrow indicating the direction of desirability.

CASE III. If one test correlates positively with a criterion and the other negatively, but both are large, ordinarily the tests will correlate negatively with each other. In this case it is desirable that the tests shall correlate as *highly* as possible with each other. The yields from the various combinations are shown in Table 83. The human organism is so constituted, however, that in this case reliable correlations of any size above zero are rarely obtained.

TABLE 82

SHOWING FOR A PAIR OF TESTS HAVING HIGH NEGATIVE CRITERION CORRELATIONS, THE CORRELATION YIELD RESULTING FROM COMBINING THEM UNDER VARYING ASSUMPTIONS AS TO THE CORRELATION WITH EACH OTHER

CRITERION CORRELATIONS, CASE II	VARYING TEST CORRELATIONS (r_{12})	DIRECTION OF DESIRABILITY OF TEST CORRELATIONS	CORRESPONDING YIELD OF COMBINED TESTS (R)
$r_{01} = -.50$ $r_{02} = -.40$	+.70		.501
	+.40		.546
	+.20		.588
	.00		.640
	-.05		.658
	-.10		.673

TABLE 83

SHOWING FOR A PAIR OF TESTS, ONE OF WHICH SHOWS A HIGH POSITIVE CRITERION CORRELATION AND THE OTHER A HIGH NEGATIVE CRITERION CORRELATION, THE CORRELATION YIELD RESULTING FROM COMBINING THEM UNDER VARYING ASSUMPTIONS AS TO THE CORRELATION WITH EACH OTHER

CRITERION CORRELATIONS, CASE III	VARYING TEST CORRELATIONS (r_{12})	DIRECTION OF DESIRABILITY OF TEST CORRELATIONS	CORRESPONDING YIELD OF COMBINED TEST (R)
$r_{01} = \pm .50$ $r_{02} = \mp .40$	+.10		.673
	+.05		.658
	.00		.640
	-.20		.588
	-.40		.546
	-.70		.501

TABLE 84

SHOWING FOR A PAIR OF TESTS ONE OF WHICH SHOWS A HIGH CRITERION CORRELATION REGARDLESS OF SIGN, AND THE OTHER A ZERO CRITERION CORRELATION, THE CORRELATION YIELD RESULTING FROM COMBINING THEM UNDER VARYING ASSUMPTIONS AS TO THE CORRELATION WITH EACH OTHER

CRITERION CORRELATIONS, CASE IV	VARYING TEST CORRELATIONS (r_{12})	DIRECTION OF DESIRABILITY OF TEST CORRELATIONS	CORRESPONDING YIELD OF COMBINED TESTS (R)
$r_{01} = \pm .50$ $r_{02} = .00$	+.70	↑ ↓	.70
	+.40		.545
	+.20		.511
	.00		.50
	-.20		.511
	-.40		.545
	-.70		.70

CASE IV. If the correlation between the criterion and one test is large and that between the criterion and the other test is zero, it is desirable that the correlation between the tests themselves shall be as large as possible regardless of sign. The situation is shown in Table 84. Since Case IV is not suggested by the general maxim given above, it should be specially noted. It shows clearly that the mere fact of a test having a correlation with the criterion which would make it of negligible value if standing by itself, does not necessarily make it so when considered as a unit in a battery.

APPLICATION OF PRINCIPLES OF FINAL TEST CHOICE TO THE FREE-HAND-DRAWING PROJECT

For purposes of examination with a view to the choice of the tests for the final battery, it is usually best to arrange the r 's in a special form such as shown in Table 85. We may now proceed to the examination of this table and the choice of the final battery for free-hand-drawing aptitude.

TABLE 85

SHOWING THE CORRELATION COEFFICIENTS OF THE FREE-HAND-DRAWING PROJECT, ARRANGED FOR CONVENIENT EXAMINATION FOR THE PURPOSE OF MAKING THE CHOICE OF THE TESTS FOR THE FINAL BATTERY

VARIABLE NUMBERS	VARIABLE NUMBERS								
	0	1	2	3	4	5	6	7	8
1	+.229								
2	+.208	+.169							
3	+.049	+.053	-.235						
4	+.002	+.100	-.008	+.096					
5	-.259	-.041	-.064	+.036	-.007				
6	+.415	+.003	+.402	-.052	-.129	-.105			
7	+.068	+.152	+.033	-.085	+.080	-.130	+.147		
8	-.321	-.020	-.154	+.085	+.091	+.130	-.120	+.044	
9	-.344	+.053	-.058	+.085	-.115	+.393	-.190	+.016	+.015

The first thing is to examine the first column of correlations, which contains the criterion coefficients. This reveals a rather gratifying situation. Six of the nine tests show correlations larger than .20, which is a higher percentage than is usually obtained in aptitude projects of this nature. These six tests satisfy the first condition of our general maxim stated above, that the correlation between the tests and the criterion should be large.

The next thing is to examine the correlations among these six tests to see whether the coefficients are sufficiently small to satisfy the second condition of our general maxim. An inspection shows that, for the most part, here also the situation is unusually satisfactory. It will be remembered that these tests were especially chosen with a view to having them overlap as little as possible. Passing now to the detailed consideration of the intercorrelations, we find that Test 1 is distinctly satisfactory. The only correlation at all suspicious is r_{12} , which is $+.169$. Test 8 also is found to have unusually

small correlations with the remaining promising tests. These two tests easily find a place in the battery.

Passing to Test 2, we find the situation not so satisfactory. The suspicious correlation with Test 1 has already been noted. The decisive factor, however, is that r_{26} is $+.402$. This is decidedly high, especially where criterion correlations as low as between $.20$ and $.40$ are involved. This high correlation is probably due to the fact that both are verbal tests. One of the two tests evidently must be discarded. This brings us to the consideration of the merits of Test 6. It turns out that the criterion correlation of this test is decidedly the best of the lot, being $+.415$. Moreover, it correlates relatively low with all of the other promising tests. It will therefore make an ideal core around which to build the remainder of the final battery. Test 2 is accordingly rejected.

Passing to the consideration of Test 5, we find the inter-correlations with all of the other promising tests to be distinctly satisfactory, with the exception of r_{59} , which is $+.393$. This raises the question as to the promise of Test 9. Since the criterion correlation of this test is next to the best, it would be unwise to discard it even though it correlates with Test 6 to the extent of $-.19$. A careful reëxamination of the correlations of Test 5 with the other promising tests shows that with the exception of r_{59} the other coefficients are all very satisfactory. It will evidently make a useful member of the battery — distinctly more useful than Test 2. It will therefore be retained.

Lastly, the tests showing small or zero correlations with the criterion are examined (Case IV, page 453) to see if any of them have high correlations with other tests which correlate highly with the criterion. Tests 3, 4, and 7 are in this class. However, the largest coefficient in this series is r_{14} , which is only $+.152$. This is much too small to make the test sig-

nificant. Indeed, r_{01} is so small as to preclude anything of value in this direction at the outset.

As a result of our analysis, then, we make as our choice for the final battery the following tests:

Test 1 — Memory for Design

Test 5 — Reproduction of Angles

Test 6 — Opposites

Test 8 — Circle Completion

Test 9 — Extension of Line

CHAPTER FOURTEEN

COMBINING THE TESTS TO SECURE THE MAXIMUM FORECASTING EFFICIENCY

WITH the tests of our final battery chosen, the next task is to combine them in such a way as to yield the best possible aptitude prediction. The first step is to determine statistically the correct weight to be given to each test.

IF TESTS ARE NOT WEIGHTED, THEY WEIGHT THEMSELVES

It is important to observe that the tests of a battery are always weighted — either well or badly. If the psychologist does not weight them, they weight themselves. This natural weighting takes place according to the sizes of the respective standard deviations of the tests involved. For example, if the scores of two tests are combined by simple addition and the S.D. of one is 10 while that of the other is 5, the first will have twice the influence of the second on the variability of the total score. This means that, if left to themselves, tests combine according to a purely accidental system. But since most test batteries are barely within the limits of practical usefulness, the margin of difference in forecasting efficiency between accidentally *vs.* scientifically weighted tests is likely to be decisive.

A SIMPLE METHOD OF SCIENTIFIC WEIGHTING

A number of methods are now available by which the tests of a battery may be weighted so as to give the maximum prognostic yield. The basic method was originally devised by G. U. Yule, the British statistician. Unfortunately, his procedure is so complicated and laborious that many psychologists have hesitated to use it. The procedure presently to be described is due largely to Tolley and Ezekiel (93). This method gives exactly the same values as that of Yule, yet is

comparatively simple to learn and requires for computation but a fraction of the time formerly consumed. In addition it makes possible a continuous checking system throughout the computation. The method involves merely the writing out of a number of simple normal equations from data already available. These equations, when solved, give the desired weights.

PRELIMINARY ARRANGEMENT OF DATA

The data required for the writing of the normal equations are the moments described in the preceding chapter. They will be recalled as intermediate stages in the computation of the correlation coefficients. The moments of the miniature problem may be seen in their original setting in Tables 75 and 76.

If the moments involved in a given project are arranged systematically in the skeleton form shown in Table 86, the normal equations may be written out with ease. First record in the column headed "Number of Test" the numbers of the tests chosen to constitute the final battery. Then record the same numbers (and in the same order, but from left to right) as subscripts to the horizontal row of W 's. A zero is always placed at the top of the column headed "Criterion Moments," since zero is always the number of the criterion. There is no W here because these moments are not weighted.

Then the subscripts for the various moments may be recorded in the various squares of the table by combining the number shown at the head of the column with the number at the left of the row (see Table 87). Thus the subscript of the moment p_{01} is so numbered because it is in the column 0 and in the row 1. Moment p_{33} is so numbered because it is in column 3 and row 3. Many of the moments such as p_{35} appear twice. The table, of course, is made large or small according to the number of tests being weighted.

TABLE 86

SHOWING THE BLANK FORM FOR ENTERING THE VARIOUS MOMENTS
PREPARATORY FOR WRITING THE NORMAL EQUATIONS

[illegible]

TABLE 87

SHOWING THE GENERAL METHOD OF WRITING THE SUBSCRIPTS FOR THE MOMENTS IN THE PRELIMINARY ARRANGEMENT OF DATA PREPARATORY TO WRITING THE NORMAL EQUATIONS. EACH MOMENT RECEIVES AS ONE OF ITS SUBSCRIPTS THE NUMBER OF ITS ROW, AND AS THE OTHER SUBSCRIPT THE NUMBER OF ITS COLUMN.

[illegible]

TABLE 88

SHOWING THE ACTUAL ARRANGEMENT OF THE MOMENTS OF THE MINIATURE PROBLEM READY FOR THE WRITING OF THE NORMAL EQUATIONS

NUMBER OF TESTS	CRITERION MOMENTS (NOT WEIGHTED) (0)	TEST MOMENTS (WEIGHTED)					
		W_1	W_2	W_3	W	W	W
1	$p_{01} = +5$	$p_{11} = +4$	$p_{12} = +1$	$p_{13} = +3$			
2	$p_{02} = -1$	$p_{12} = +1$	$p_{22} = +4$	$p_{23} = -6$			
3	$p_{03} = +9$	$p_{13} = +3$	$p_{23} = -6$	$p_{33} = +16$			

Lastly, the numerical moments corresponding to the subscripts already written into the form are recorded in the respective squares. These will be found in the next-to-last column of the work sheets for the standard deviations and the correlation coefficients, respectively. These moments should be transferred to the new table with special care, particularly as to signs. The moments of the miniature problem are shown properly arranged in Table 88.

HOW TO WRITE THE NORMAL EQUATIONS

We are now ready to write the normal equations. Each row of the entries in the table makes up one equation. The general form of the first equation is:

$$p_{11}W_1 + p_{12}W_2 + p_{13}W_3 + p_{14}W_4 + p_{15}W_5 \text{ etc.} = p_{01} \quad (21)$$

This means that, skipping the criterion moment, we multiply the first test moment by its weight (W_1) shown at the top of

the column, the second test moment by its weight (W_2), and so on until we have multiplied all the moments by their respective weights. These are connected by placing before each product the sign of the moment involved. Lastly, the whole is made equal to the criterion moment, which has no weight. The equations of the miniature problem, written from Table 88, are as follows:

$$4 W_1 + 1 W_2 + 3 W_3 = 5$$

$$1 W_1 + 4 W_2 - 6 W_3 = -1$$

$$3 W_1 - 6 W_2 + 16 W_3 = 9$$

The values of the W 's (weights) may now be found by successively eliminating one W after the other by the methods of elementary algebra for the solution of simultaneous equations.

A SYSTEM THAT INSURES ACCURACY IN THE COMPUTATION OF WEIGHTS

Just as in the computation of the correlation coefficients and of the moments upon which the present equations are based, it is very desirable here also that a check be provided which will detect any inaccuracy in the computations leading to the weights. An undetected error at this point is likely to ruin an entire project. Fortunately there is a method by means of which the accuracy of the arithmetic may be tested at any point throughout the process of elimination. This is especially important where four or more tests are being weighted, because otherwise an error may not be discovered until after several hours of labor have been wasted. The checking method is essentially that of Gauss. It is based on a special check column placed at a convenient distance to the right of the equations.

The items at the top of the check column are made up of the algebraic sums of the coefficients of the W 's and the

constant term (moments) contained in the respective equations, the sign of equality being ignored. Thus the algebraic sum of the coefficients and constant term of the first equation of the miniature problem (page 463) is $+13$, that of the second is -2 , and so on.

After the *original* normal equations have been added up, however, the later entries in the check column are derived in an entirely different manner. They are derived in the same manner that entries in the successive equations themselves are derived. In general, whenever an operation is performed on an equation the same operation is also performed on the check number of that equation. For example, in the solution of the miniature problem given below (page 463), equation No. 4 is derived by multiplying the members of equation No. 2 by -4 . Accordingly, the check entry for No. 2 (which is a -2) is also multiplied by -4 . This makes the check entry for equation No. 4 a $+8$. Later, equations Nos. 1 and 4 are added. Consequently, the corresponding check numbers of these equations are also added. The check number of equation No. 1 ($+8$) and the check number of equation No. 4 ($+13$) make a total of $+21$, which is the check entry of equation No. 6.

But suppose that at this point it were desired to test equation No. 6 for accuracy. It is merely necessary to find the algebraic sum of the coefficients of equation No. 6; i.e., $27 + 9 - 15 = 21$. This computation agrees with the check entry for this equation, which, as we have seen, was obtained in an entirely different manner. The fact that the two methods of securing this check number agree in the result obtained indicates that all the arithmetic up to this point is correct. This test for arithmetical accuracy should be applied frequently throughout the computation for the weights. In particular it should always be applied to the last equation of the series.

AN ILLUSTRATIVE EXAMPLE OF DETERMINING WEIGHTS,
FOUR VARIABLES

In order to illustrate both the mode of solution of the equations and the method of checking just described, the computation of the weights for the tests of the miniature problem will be given in detail :

EQUATIONS	CHECK COLUMN	NUMBER OF EQUATION
$4 W_1 + 1 W_2 + 3 W_3 = + 5$	+ 13	(1)
$1 W_1 + 4 W_2 - 6 W_3 = - 1$	- 2	(2)
$3 W_1 - 6 W_2 + 16 W_3 = + 9$	+ 22	(3)
Multiplying equation (2) by - 4 and equation (3) by - 1.333 (i.e., $\frac{4}{3}$) so as to cancel out W_1 when added to equation (1), we have:		
$- 4 W_1 - 16 W_2 + 24 W_3 = + 4$	+ 8	(4)
$- 4 W_1 + 8 W_2 - 21.333 W_3 = - 12$	- 29.333	(5)
Adding (4) to (1), then (5) to (1):		
$- 15 W_2 + 27 W_3 = 9$	+ 21	(6)
$9 W_2 - 18.333 W_3 = - 7$	- 16.333	(7)
Multiplying (7) by + 1.667 so as to cancel out W_2 when added to (6):		
$+ 15 W_2 - 30.555 W_3 = - 11.667$	- 27.222	(8)
Adding (6) and (8):		
$- 3.555 W_3 = - 2.667$	- 6.222	(9)
$W_3 = \frac{- 2.667}{- 3.555} = + .75$		
Substituting the value of W_3 in (6):		
$- 15 W_2 + 27 \times .75 = 9$		
$- 15 W_2 = - 11.25$		
$W_2 = \frac{- 11.25}{- 15} = + .75$		
Substituting the values of W_2 and W_3 in (1):		
$4 W_1 + 1 \times .75 + 3 \times .75 = + 5$		
$4 W_1 = 2$		
$W_1 = \frac{2}{4} = + .5$		

We now have the weights to the three tests, which we set out to secure.

The method of checking the detailed steps of the elimination process has been given above. After finding all the weights, it is well (particularly for the novice) also to check the entire procedure. This may be done by substituting all the W 's in one of the original equations which has not yet been used for substitution purposes. For example, in the miniature problem considered above all the W 's may be checked by being substituted in equation (2) :

$$\begin{aligned} 1 \times .5 + 4 \times .75 - 6 \times .75 &= -1 \\ .5 + 3 - 4.5 &= -1 \\ -1 &= -1 \end{aligned}$$

If any error has been made either in the substitution operations or indeed anywhere in the solution of the simultaneous equations, the equation usually will not hold when such a substitution is made. In actual problems there will always be some dropped decimals which will prevent an absolute agreement such as shown above, but the difference should not be great.

CONSTRUCTION OF PREDICTION FORMULA, FOUR VARIABLES

Having the weights and the means of the various tests of a battery, we may now proceed to the construction of the prediction formula. The standard formula for this purpose is the multiple-regression equation. The use of this most valuable instrument has been discussed incidentally in previous chapters, notably pages 179 ff. We have now to consider the method of preparing it for use. The general form of the multiple-regression equation most convenient in aptitude work is :

$$\begin{aligned} \bar{X}_0 &= W_1X_1 + W_2X_2 + W_3X_3 + W_4X_4 + W_5X_5, \text{ etc.} \\ &+ M_0 - W_1M_1 - W_2M_2 - W_3M_3 - W_4M_4 - W_5M_5, \text{ etc. (22)} \end{aligned}$$

In this formula the W 's are the weights of the tests, the M 's are the means of the corresponding tests and criterion series,

and the X 's are the test scores of any particular person for whom an aptitude estimate is desired. The $\bar{X}_0$ represents the aptitude forecast of a given person to be secured by the use of the formula. It is distinguished from the criterion of the aptitude itself (X_0) by having a bar above the X .

To prepare a prediction formula for use with a particular battery of tests it is merely necessary to substitute in the general formula given above the means and weights of the respective tests, then combine and collect terms. The elements in the general equation having subscripts corresponding to tests not found in the battery simply drop out. In the miniature problem already considered, the substitution appears as follows:

$$\bar{X}_0 = .5 X_1 + .75 X_2 + .75 X_3 + 7 - .5 \times 8 - .75 \times 3 - .75 \times 7$$

The values in the general formula with subscripts beyond 3 drop out because there are no tests with such numbers in this battery. Performing the indicated multiplications:

$$\bar{X}_0 = .5 X_1 + .75 X_2 + .75 X_3 + 7 - 4 - 2.25 - 5.25$$

Uniting terms, we have as a prediction formula:

$$\bar{X}_0 = .5 X_1 + .75 X_2 + .75 X_3 - 4.5$$

If no error has been made in its preparation, the formula is now ready for use.

THE FORMULA IN USE

Suppose now a person, John Doe, of unknown aptitude, should be tested by means of this battery, and should make scores of 7, 4, and 5 on the respective tests. According to the above notation and for this particular subject:

$$X_1 = 7$$

$$X_2 = 4$$

$$X_3 = 5$$

Substituting these values in the prediction formula, we have:

$$\bar{X}_0 = .5 \times 7 + .75 \times 4 + .75 \times 5 - 4.5$$

$$\bar{X}_0 = 3.5 + 3 + 3.75 - 4.5$$

$$\bar{X}_0 = 5.75$$

Thus 5.75 is found to be the forecast of the potential aptitude of the subject. Since the mean ability in this aptitude is 7, this forecast indicates that John Doe probably will turn out to be somewhat below the average, in case he should attempt to learn the occupation in question.

HOW TO FIND THE CORRELATION YIELD (R) OF A BATTERY BY FORMULA

The multiple-regression equation gives the closest estimate it is possible to secure from any particular battery of tests.¹ If the battery contains test units of considerable prognostic value, the equation will make the most of them. If, however, the test units are of little or no value, then the forecasts made by the equation will be without significance. This will be the fault not of the equation but of the tests. If the correlation yield is less than .50 or so, the battery will not be worth using (pages 273 ff.). It accordingly becomes a matter of considerable interest as well as of practical importance to know the forecasting power of a battery. This is indicated by the correlation between (A) the predicted or estimated aptitude and (B) the original aptitude criterion. This value is usually represented by R and is called the *coefficient of multiple correlation*. It may be determined by two independent methods. This fact not only enables each method to serve as a check on the correctness of the other but, fortunately, *between them they check the accuracy of the*

¹Strictly speaking, this holds only where the relations are all linear and then exactly only for the particular data, taken as a whole, on which the equation was based. Fortunately, most relations in aptitude work are approximately linear.

work from the very beginning of the preliminary correlation work sheet.

The simplest of the two methods of determining the coefficient of multiple correlation is by means of the formula :

$$R = \sqrt{\frac{p_{01}W_1 + p_{02}W_2 + p_{03}W_3 + p_{04}W_4 + p_{05}W_5, \text{ etc.}}{p_{00}}} \quad (23)$$

An inspection of this formula reveals it to be simply the square root of the quotient of the sum of the product moments by the corresponding weights, divided by the square moment of the criterion. This will become clear if we substitute in the formula the appropriate values of the miniature problem the weights of which were found above. The moments are given in Tables 75 and 76.

$$R = \sqrt{\frac{5 \times .5 - 1 \times .75 + 9 \times .75}{16}}$$

Solving by easy stages :

$$\begin{aligned} R &= \sqrt{\frac{2.5 - .75 + 6.75}{16}} \\ &= \sqrt{\frac{8.50}{16}} \\ &= \sqrt{.5312} \\ &= .7288 \end{aligned}$$

The multiple-correlation coefficient of the miniature problem is therefore .7288. This corresponds to a forecasting efficiency (page 273) of 31.5 per cent.

THE CORRELATION YIELD BY DIRECT COMPUTATION

The second method of determining R is first to substitute in the prediction formula the test scores which were employed in evolving it. The resulting criterion estimates may then be correlated with the original criterion scores. The coeffi-

cient obtained in this way also is designated as R . If the computations have been done with sufficient accuracy, the two values of R will agree to the third or fourth decimal place. Accordingly, if we substitute the test scores of our miniature problem (Table 74) one subject at a time, we obtain the five "forecasts" or estimates shown in the first column of Table 89. The second column gives the original criterion scores, taken from Table 74.

TABLE 89

$\bar{X}_0$		X_0
	10.5	13
	7.0	2
	4.0	4
	3.5	6
	10.0	10
Sum	35.0	35
Mean	7	7
S.D.	2.915	4

$$R = .7288$$

Carrying out in the ordinary manner the various computations for the product-moment correlation coefficient, we find that the two columns of Table 89 correlate to the extent of .7288. This is exactly the same as obtained by formula. Thus the two R 's agree to the fourth decimal place. This confirms the accuracy of the computations from the very beginning of the project. No prediction formula should ever be used which has not been checked for accuracy in this manner.

LIMITATIONS ON THE SIGNIFICANCE OF R

The computations just given have shown that the coefficient of multiple correlation (R) as found by the formula agrees exactly with the correlation between the original and the

estimated criterion scores *when the estimates are based on the data from which the forecasting formula was derived*. It is important to observe, however, that this does not hold exactly, but only approximately, when the battery is applied to a new group of subjects. There is a natural tendency for the equation to fit, in general, the group from which it was derived somewhat better than any other group. In other words, there is a tendency for the battery to yield somewhat smaller correlations in actual use than the R given by the formula. This shrinkage is larger the *more tests* in the battery and the *fewer subjects* in the original experimental group.

Fortunately, there is another tendency which compensates more or less for the shrinkage just mentioned. This is due to the fact that the correlations found by ordinary methods between aptitude estimates and aptitude criteria are never as great as the true correlations. This, in turn, is due to the fact that aptitude criteria are practically never perfect measures of the true aptitudes. This reduction in the size of the correlation coefficient below its true value is known as attenuation (pages 231 ff.). If the criterion score has been obtained from two independent sources from which a reliability coefficient may be obtained, it is possible to estimate pretty accurately what the true correlation is (pages 237 ff.). It will be found invariably to be larger than indicated by R . In most cases, therefore, the net result is for the two tendencies about to neutralize each other.

If it were desired to secure an R free from both these distortions, it could be done by administering the tests to an entirely new group of subjects (making the aptitude estimates by means of the forecasting formula derived from the first group) and then correcting for attenuation the correlation found to exist between the predicted and the actual criterion scores of the second group.

CRITERION ESTIMATES MADE BY MEANS OF REGRESSION
EQUATION HAVE NARROWED DISPERSION

One important peculiarity of forecasts made by means of the multiple-regression equation is that the range, or dispersion, of the estimates is less than that of the original criterion scores. This is shown in the case of the estimates for the miniature problem in Table 89. The standard deviation of the original criterion is 4, whereas that for the estimates made by means of the equation is only 2.915. The range of the estimated criterion is dependent upon the size of R . In fact, it can be determined before the forecasts are made, simply by multiplying the standard deviation of the original criterion by R . For example, in the miniature problem, σ is 4 and R is .7288. Accordingly

$$4 \times .7288 = 2.915$$

This yields exactly the same value (Table 89) as was obtained by computation from the forecasts themselves.

This shrinkage in the dispersion of estimates made by means of a multiple-regression equation is inherent and significant. It in no sense implies inaccuracy or defect on the part of the regression equation. Indeed, it is necessary in order that the error of estimate may be reduced to a minimum. But even so it does introduce a certain amount of ambiguity from the point of view of the interpretation of aptitude forecasts. It is somewhat as if the estimates were being made on a new and unfamiliar scale. The various ranges of aptitude forecasts for typical values of R are shown in Table 90. Where the correlation is perfect, there is no distortion. It is interesting to note, on the other hand, that if a test battery has no forecasting value whatever, all subjects are estimated by its equation as having exactly the same aptitude, regardless of their test scores; all alike are esti-

TABLE 90

SHOWING THE LIMITING FORECAST VALUES OF THE MULTIPLE-REGRESSION EQUATION FOR VARIOUS VALUES OF R , IN TERMS OF SCHOOL MARKS

R	FORECAST VALUES		
	Highest Possible Mark	Mean Mark	Lowest Possible Mark
1.00	100	81	62
.80	96	81	66
.60	92	81	70
.40	89	81	73
.20	84	81	77
.00	81	81	81

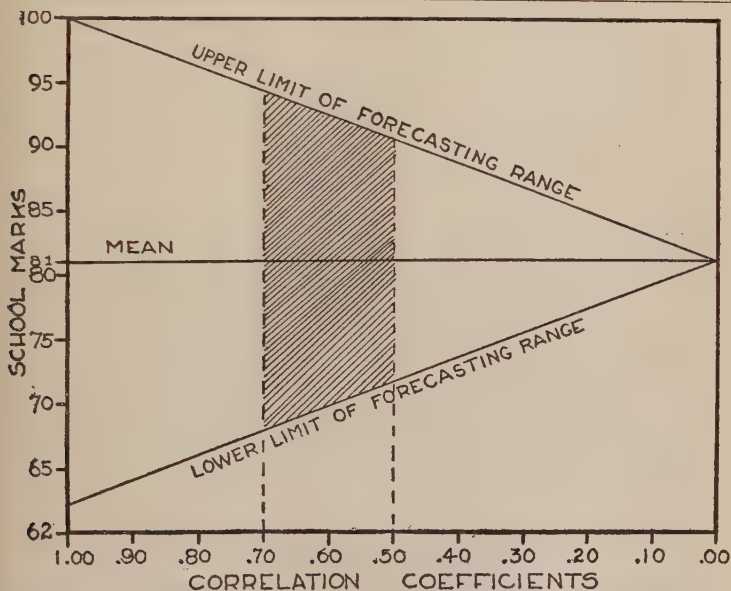


FIG. 59. Showing the different ranges in aptitude forecasts made by the multiple-regression equation for different values of R . The shaded area shows the forecasting range for aptitude batteries within the zone of practical utility.

mated at the mean. This is shown graphically by Figure 59. Fortunately, from this point of view, the correlation zone of test batteries of practical usefulness is fairly narrow. As a result, the difference in the forecasting ranges of practical test batteries is not great. It ranges closely around 60 per cent that of the original criterion.

PRACTICAL DIFFICULTIES ARISING FROM SHRINKAGE IN DISPERSION

As a result of the peculiarity just mentioned, persons inexperienced in the use of the regression equation may be misled by the forecasts. For example, if an equation with R of 60 were being used to forecast school grades on some unknown individuals, there would be found no persons with excellent or with non-passing grade predictions, and very few with good or poor grade predictions. It might easily be mistakenly assumed that no very superior or very inferior individuals existed in the new group. If it were desired, for administrative purposes, to select a certain per cent of individuals at the upper or lower extremes of aptitude, a distinctly different figure would need to be chosen as a dividing point, in the case of regression forecasts, from what would be required if the original criterion were to be used. For example, approximately 6 per cent of all subjects fall below 70 on the theoretical marking system. But with the regression equation (where R is .50) a forecast of 75.5 must be selected in order to cut off a similar number.

A PREDICTION FORMULA WHICH DOES NOT DISTORT THE CRITERION DISPERSION

Fortunately, it is an easy matter to reconstruct the regression equation so that estimated criterion scores will have exactly the same range as the original criterion scores. This

the criterion estimate without in the least disturbing the correlation or the mean is easily shown by making the estimates anew by means of it. These are shown in column $\bar{X}_0$ of Table 91.

TABLE 91

	$\bar{X}_0$	X_0
	11.80	13
	7.00	2
	2.88	4
	2.20	6
	11.12	10
Sum	35.00	35
Mean	7	7
S.D.	4.000 +	4

$$R = .7288$$

Here the σ of the estimates is seen to agree exactly with that of the original criterion without in any way disturbing the other statistical values.

A FORMULA FOR FORECASTING ON ANY DESIRED SCALE OTHER THAN THE CRITERION

It sometimes happens that it is desired to forecast an aptitude in terms of an entirely different system or scale from that of the original criterion. It may be that the original unit of criterion measurement is undesirable for purposes of forecasting the aptitude in question. For example, the original criterion score of a scholastic aptitude may have been obtained in terms of the number of items correctly performed on an achievement test. This might serve very well from the point of view of adequate measurement, but would not be a satisfactory practical forecasting scale because people are unfamiliar with it. The natural prediction unit in such a

case is obviously that of the ordinary school marking system. Accordingly, where the original criterion score is not what is desired for the criterion estimate, it may be transmuted readily into any other system that may be desired. The method of transmutation has already been described above (pages 397 ff.). If this is to be done, it should be performed before the correlation computation begins.

Frequently, however, it is desired to change the prediction scale of a forecasting formula *after* the formula has been completely worked out. Various circumstances may bring this about. Often some peculiarity of the original criterion score, which makes it undesirable for forecasting purposes, has not been realized until the formula was nearly or quite completed. Such formulæ may be transformed in a few moments into equations which will make forecasts in terms of any system or scale that may be desired (36). Indeed, it is usually simpler to make the transformation after the completion of the equation than to transmute the criterion scores before beginning the computations for the formula.

In this transformation of the multiple-regression equation there is involved not only the dispersion of the criterion forecasts, but also their means. Both are provided for in the following general formula:

$$\bar{X}'_0 = CW_1X_1 + CW_2X_2 + CW_3X_3 + CW_4X_4 + CW_5X_5, \text{ etc.} \\ + M'_0 - CW_1M_1 - CW_2M_2 - CW_3M_3 - CW_4M_4, \text{ etc.} \quad (25)$$

where $\bar{X}'_0$ is the new criterion estimate, M'_0 is the mean of the new criterion system, and

$$C = \frac{\sigma'_0}{R\sigma_0}$$

The σ_0 is the standard deviation of the original criterion score, σ'_0 is the standard deviation desired for the predicted criterion, and R is the multiple-correlation coefficient.

Perhaps this may be made clearer by an example. Suppose it were desired to transform the multiple-regression equation of the miniature problem into a forecasting formula which would predict in terms of school marks with a mean of 81 and a S.D. of 7. In this case, then :

$$\begin{aligned} R &= .7288 \\ \sigma_0 &= 4 \\ \sigma_0' &= 7 \\ M_0' &= 81 \end{aligned}$$

Accordingly

$$\begin{aligned} C &= \frac{7}{.7288 \times 4} \\ &= 2.4012 \end{aligned}$$

Substituting in the general equation, we have :

$$\begin{aligned} \bar{X}_0' &= 2.4 \times .5 X_1 + 2.4 \times .75 X_2 + 2.4 \times .75 X_3 + 81 \\ &\quad - 2.4 \times .5 \times 8 - 3.4 \times 7.5 \times 3 - 2.4 \times .75 \times 7 \end{aligned}$$

Performing the indicated operations and uniting terms :

$$\bar{X}_0 = 1.2 X_1 + 1.8 X_2 + 1.8 X_3 + 53.38$$

That this equation accomplishes the desired end is easily seen by making the criterion estimates for the miniature problem by means of it. These estimates are given in column $\bar{X}_0'$ of Table 92. These computations show that the transformed equation makes estimates strictly in terms of the school marking system, yet without in the least disturbing the maximum correlation.

If it should be desired to forecast in terms of a scale different from that of the original criterion, yet by means of a true regression equation, the procedure is exactly the same as that just described, except that in the present case :

$$C = \frac{\sigma_0'}{\sigma_0}$$

TABLE 92

$\bar{X}_0'$		X_0
89.40		13
81.00		2
73.80		4
72.60		6
88.20		10
Sum	405.00	35
Mean	81	7
S.D.	7.00 —	4

$R = .7288$

If it should be desired to predict with a different mean from the original criterion, but with the characteristic dispersion produced by the true regression equation undisturbed, the general formula becomes :

$$X_0 = W_1X_1 + W_2X_2 + W_3X_3 + W_4X_4 + W_5X_5 \text{ etc.}$$

$$+ M_0' - W_1M_1 - W_2M_2 - W_3M_3 - W_4M_4 - W_5M_5$$

etc. (26)

This is the same as the regular regression equation, except that M_0 has been replaced by M_0' , which is the mean desired for the predicted series.

THE FREE-HAND-DRAWING PROJECT

We may now proceed to apply the methods described above for the construction of forecasting formulæ, to the concrete problem of free-hand-drawing aptitude. The first thing is to draw up the form for the table of moments. The numerical values to be entered into this table are found in Tables 78 and 79, respectively. The moments thus assembled are shown in Table 93. In order to save labor in computation, the smaller decimals have been dropped. It has been neces-

TABLE 93

SHOWING THE TABLE OF MOMENTS FOR THE TESTS OF THE FINAL BATTERY FOR FREE-HAND DRAWING

NUMBER OF TEST	CRITERION MOMENTS (NOT WEIGHTED) (0)	TEST MOMENTS (WEIGHTED)				
		W_1	W_2	W_3	W_4	W_5
1	$p_{01} = + 6.563$	$p_{11} = + 16.44$	$p_{15} = - 1.396$	$p_{16} = + .0574$	$p_{18} = - .1099$	$p_{19} = + .0440$
5	$p_{05} = - 15.18$	$p_{15} = - 1.396$	$p_{55} = + 69.11$	$p_{56} = - 3.841$	$p_{58} = + 1.446$	$p_{59} = + .6630$
6	$p_{06} = + 12.82$	$p_{16} = + .0574$	$p_{56} = - 3.841$	$p_{66} = + 19.20$	$p_{68} = - .7055$	$p_{69} = - .1693$
8	$p_{08} = - 3.031$	$p_{18} = - .1099$	$p_{58} = - 1.446$	$p_{68} = - .7055$	$p_{88} = + 1.793$	$p_{89} = + .0411$
9	$p_{09} = - .4931$	$p_{19} = + .0440$	$p_{59} = + .6630$	$p_{69} = - .1693$	$p_{89} = + .0411$	$p_{99} = + .0413$

sary, however, to retain about four figures in each case in order to secure the desired degree of accuracy in the final results.

THE NORMAL EQUATIONS AND THEIR SOLUTION

The normal equations written from this table of moments appear as the first five equations of the chart, page 479. Below these are shown in detail the computations by which are derived the weights for the five tests, together with the check column for testing the accuracy of the work as it progresses. Because of even small decimals which must be dropped constantly, these checks are not quite exact, especially near the end of the computation where the number of decimals thus dropped becomes rather large. It is quite accurate enough, however, to detect any significant error in the computation.

CHART SHOWING THE DETAILED COMPUTATIONS BY WHICH W_e WERE DERIVED THE WEIGHT FOR THE FIVE TESTS
OF THE APTITUDE BATTERY FOR FREE-HAND DRAWING

	CHECK COLUMN	EQUATION No.
$+ 11.77650 \times (2),$		(1)
$- 286.41114 \times (3),$	6.563	(2)
$+ 149.59053 \times (4),$	15.18	(3)
$- 373.63636 \times (5),$	27.3616	(4)
	.5663	(5)
	.1270	
$- 16.44 W_1 + 813.874 W_2 - 45.2335 W_3 + .0574 W_4 - 1.099 W_5 + .0440 W_6 = +$	598.2697	(6)
$- 16.44 W_1 + 1100.105 W_2 - 5499.093 W_3 + 202.063 W_4 + 48.4894 W_5 - 7836.6670$	7836.6670	(7)
$- 16.44 W_1 + 216.3079 W_2 - 105.5361 W_3 + 268.2168 W_4 + 6.1482 W_5 - 84.7131$	84.7131	(8)
$- 16.44 W_1 - 247.7209 W_2 + 63.2566 W_3 - 15.3565 W_4 - 15.4312 W_5 + 184.2401$	184.2401	(9)
$+ 812.478 W_2 - 45.1761 W_3 + 16.9189 W_4 + 7.8518 W_5 - 172.2043$	619.8682	(10)
$+ 1098.709 W_2 - 5499.0356 W_3 + 201.9531 W_4 + 48.5334 W_5 - 3665.2278$	7815.0685	(11)
$+ 214.9119 W_2 - 105.4787 W_3 + 268.1059 W_4 + 6.1922 W_5 - 446.8459$	63.1146	(12)
$- 249.1169 W_2 + 63.3140 W_3 - 15.4664 W_4 - 15.3872 W_5 + 190.8031$	25.8533	(13)
$- 812.478 W_2 + 4066.4288 W_3 - 149.3403 W_4 - 35.8895 W_5 - 2710.3627$	5779.0869	(14)
$- 812.478 W_2 + 398.7633 W_3 - 1013.5770 W_4 - 23.4097 W_5 - 1689.3054$	238.6054	(15)
$- 812.478 W_2 + 206.4942 W_3 - 50.4426 W_4 - 50.1843 W_5 - 622.2910$	84.3187	(16)
$+ 4021.2507 W_2 - 132.4214 W_3 - 28.0377 W_4 - 2538.1584$	6398.9551	(17)
$+ 353.5872 W_2 - 996.3581 W_3 - 15.5570 W_4 - 1517.1011$	858.4736	(18)
$+ 161.3181 W_2 - 33.5237 W_3 - 42.3525 W_4 - 450.0867$	535.5495	(19)
$- 4021.2507 W_2 + 1133.47135 W_3 + 176.9356 W_4 - 17253.5660$	9763.1709	(20)
$- 4021.2507 W_2 + 835.6607 W_3 + 1055.2419 W_4 - 1219.5182$	13349.8887	(21)
$+ 11202.2921 W_2 + 148.8979 W_3 - 14715.4076$	3364.2248	(22)
$+ 703.2393 W_2 + 1027.2042 W_3 - 8681.3598$	6950.9336	(23)
$- 11202.2921 W_2 - 16362.9109 W_3 + 138290.2418$	110726.3138	(24)
$- 16214.0130 W_3 - 123574.8342$	107301.0195	(25)
$11202.2921 W_2 - 1134.8253 W_3 - 14715.4076$	7.6215	
$4021.2507 W_2 + 160.5318 + 213.6888 W_3 = +$	1.2123	
$W_4 = +$	2538.1584	
$W_5 = +$	.5379	
$812.478 W_2 - 24.3002 - 20.5104 - 59.84234 W_6 = -$	172.2043	
$16.44 W_1 + .1167 + .0309 + .1332 - .3353 W_7 = +$	.0836	
$.02239 + .32111 + 10.3276 + .85526 + 1.2903 W_8 = +$	6.563	
$+ 12.82$	3.86	
$+ 12.82$	12.82	
$+ 12.82$	12.82	

THE MULTIPLE-REGRESSION EQUATION

We may now construct our multiple-regression equation. Substituting in the general formula the weights of the several tests as found in the chart (page 479), we have:

$$\bar{X}_0 = .4 X_1 - .08 X_5 + .54 X_6 - 1.2 X_8 - 7.6 X_9 + 81 - .4 \times 18.3 + .08 \times 20.4 - .54 \times 7.4 + 1.2 \times 3.6 + 7.6 \times 4.2$$

The necessary M 's for substitution are shown in Table 77. Performing the multiplications and uniting terms:

$$\bar{X}_0 = .4 X_1 - .08 X_5 + .54 X_6 - 1.21 X_8 - 7.6 X_9 + 78.87$$

This is our multiple-regression equation. It will be observed that the weights here used are much shorter numbers than those shown in the chart on page 479. The less significant figures of the various decimals have been dropped in order to make the equation more economical to use. It will be observed also that 81 was substituted for 80.9 as M_0 , since the mean desired of the predicted criterion is 81, rather than 80.9, which was the mean of the actual criterion.

THE CORRELATION YIELD (R), FREE-HAND-DRAWING
APTITUDE BATTERY

To find the correlation between the test battery thus weighted and the criterion, we substitute values already found, in Formula 23:

$$R = \sqrt{\frac{.395(6.562) - .084(-15.184) + .538(12.815) - 1.212(-3.031) - 7.62(-.473)}{49.774}}$$

$$R = \sqrt{\frac{2.592 + 1.2755 + 6.8945 + 3.6736 + 3.7567}{49.774}}$$

$$R = \sqrt{\frac{18.1923}{49.774}}$$

$$R = \sqrt{.3654}$$

$$R = .605$$

By Formula 11 this R is found to correspond to a forecasting efficiency of slightly over 20 per cent. This is far less than might be desired, but is fully as good as the average of present-day batteries.

THE FINAL CHECK ON THE ARITHMETIC

As a final check on the accuracy of the computations, the test scores of each subject are substituted in the formula and the criterion estimates made. These are shown, together with the original criterion scores, in Table 94. The correlation between the original criterion scores and these estimated criterion scores is found by direct computation to be .607. This differs from R by .002. This is evidently due to dropped decimals. The difference usually does not exceed .005.

PRACTICAL DEFECTS OF THE MULTIPLE-REGRESSION EQUATION AS ORDINARILY USED

We must now consider certain disadvantages encountered in the use of multiple-regression equations and allied forecasting formulæ. While these formulæ make the most effective possible use of whatever virtues may exist in a battery of tests, they are likely to require somewhat more labor, as ordinarily employed, than the rough-and-ready methods frequently met with for combining test scores. The main source of the difficulty lies in the multiplication of the test scores by rather large and clumsy decimals, with the attendant necessity of later adding relatively large numbers, the latter also involving decimals. The amount of adding is increased by the presence of the constant term in the formula. The situation is aggravated still further by the fact that many of the weights may be negative, which complicates matters by requiring first a summation of the positive and the negative values separately, after which it is necessary to find the alge-

TABLE 94

SHOWING THE ESTIMATED CRITERION SCORES BY THE SIDE
OF THE ORIGINAL CRITERION SCORES

SUBJECT No.	ORIGINAL CRITERION Scores (X_0)	ESTIMATED CRITERION Scores ($\bar{X}_0$)
1	93	83.0
2	72	76.3
3	90	89.3
4	79	84.6
5	79	76.4
6	89	82.8
7	90	82.4
8	91	80.7
9	82	74.1
10	80	81.8
11	80	77.7
12	80	82.3
13	89	81.4
14	83	77.0
15	76	77.7
16	87	84.0
17	91	86.6
18	71	73.8
19	65	74.0
20	81	74.1
21	74	83.8
22	68	79.0
23	85	85.6
24	85	82.6
25	82	80.4
26	69	74.6
27	75	79.4
28	83	81.3
29	97	85.2
30	73	81.3
31	78	79.4
32	82	86.7
33	87	85.3
34	89	82.6
35	80	88.9
36	80	79.5
37	71	79.1
38	73	75.8
39	77	77.1
40	73	77.5
41	87	87.8
42	79	81.4
43	85	89.2
44	80	76.6
45	78	75.6
46	77	82.9
47	80	79.8
48	88	85.3
Mean	80.9	80.9
S.D.	7.06	4.26

$$R_{X_0\bar{X}_0} = .607$$

braic difference between the two sums. This last difficulty, however, will be encountered in any method if the battery includes tests correlating both positively and negatively with the criterion, and cannot be charged to the prediction formula.

The difficulty of multiplication may be reduced greatly by dropping off the less significant figures of the decimals making up the weights. It will be recalled that this was done above. It is probably rare where more than two digits need to be retained for the practical value of the battery to be realized in the forecast.

INCONVENIENCES OF SCIENTIFIC WEIGHTING ELIMINATED
BY A FACILITATING TABLE

If a battery is to be used very extensively, all of the multiplication, all of the subtraction, and the addition of the constant term should be eliminated. This may be done very effectively and conveniently by a suitably prepared table. The principles and methods of the preparation and use of such a table may be explained by means of a concrete example. In the case of the free-hand-drawing aptitude project we have in this respect about as difficult a case to handle as is likely to be encountered. Three of the five tests have negative weights, while the magnitude and range of the various gross test scores vary extensively. It will accordingly serve as an excellent sample to demonstrate the possibilities of the method.

The facilitating table in question is shown as Table 95. Its use is as follows: Suppose a person is tested with this battery under standard conditions and earns gross test scores as follows:

TEST No.	GROSS SCORES	WEIGHTED SCORES ACCORDING TO FACILITATING TABLE
1	17	53.7
5	25	6.0
6	16	8.6
8	2.6	8.9
9	.65	7.0
		84.2 = Forecast

The weighted scores equivalent to the respective gross scores are found in the facilitating table in a few seconds and are entered opposite the gross scores, as shown. This column, when added, gives the aptitude forecast directly. In the present case it is 84.2.

HOW TO CONSTRUCT THE FACILITATING TABLE

The method of constructing the table can best be explained by first making the same forecast by substituting in the prediction formula in the ordinary manner:

$$\bar{X}_0 = .4 \times 17 - .08 \times 25 + .54 \times 16 - 1.2 \times 2.6 - 7.62 \times .65 + 78.87$$

Performing the indicated multiplications:

$$\bar{X}_0 = 6.8 - 2.0 + 8.64 - 3.12 - 4.95 + 78.87$$

Combining positive and negative values:

$$\bar{X}_0 = 94.31 - 10.07$$

Subtracting:

$$\bar{X}_0 = 84.2$$

The values appearing in the second step given above correspond to what have been called "weighted scores" as looked up in the facilitating table. They are not identical, however. For tabular purposes the constant term, 78.87, first was shortened to 78.9. It was next found by an exami-

TABLE 95

TO FACILITATE THE MAKING OF APTITUDE ESTIMATES ON FREE-HAND
DRAWING

GROSS SCORES, TESTS 1-5-6	WEIGHTED SCORES, TEST 1	WEIGHTED SCORES, TEST 5	WEIGHTED SCORES, TEST 6	GROSS SCORES, TEST 8	WEIGHTED SCORES, TEST 8	GROSS SCORES, TEST 9	WEIGHTED SCORES, TEST 9
1	47.3	7.9	.5	.2	11.8	.05	11.6
2	47.7	7.8	1.1	.4	11.5	.10	11.2
3	48.1	7.8	1.6	.6	11.3	.15	10.9
4	48.5	7.7	2.2	.8	11.0	.20	10.5
5	48.9	7.6	2.7	1.0	10.8	.25	10.1
6	49.3	7.5	3.2	1.2	10.5	.30	9.7
7	49.7	7.4	3.8	1.4	10.3	.35	9.3
8	50.1	7.4	4.3	1.6	10.1	.40	9.0
9	50.5	7.3	4.9	1.8	9.8	.45	8.6
10	50.9	7.2	5.4	2.0	9.6	.50	8.2
11	51.3	7.1	5.9	2.2	9.3	.55	7.8
12	51.7	7.0	6.5	2.4	9.1	.60	7.4
13	52.1	7.0	7.0	2.6	8.9	.65	7.0
14	52.5	6.9	7.6	2.8	8.6	.70	6.7
15	52.9	6.8	8.1	3.0	8.4	.75	6.3
16	53.3	6.7	8.6	3.2	8.1	.80	5.9
17	53.7	6.6	9.2	3.4	7.9	.85	5.5
18	54.1	6.6	9.7	3.6	7.6	.90	5.1
19	54.5	6.5	10.3	3.8	7.4	.95	4.8
20	54.9	6.4	10.8	4.0	7.2	1.00	4.4
21	55.3	6.3	11.3	4.2	6.9	1.05	4.0
22	55.7	6.2	11.9	4.4	6.7	1.10	3.6
23	56.1	6.2	12.4	4.6	6.4	1.15	3.2
24	56.5	6.1	13.0	4.8	6.2	1.20	2.9
25	56.9	6.0	13.5	5.0	6.0	1.25	2.5
26	57.3	5.9		5.2	5.7	1.30	2.1
27	57.7	5.8		5.4	5.5		
28	58.1	5.8		5.6	5.2		
29	58.5	5.7		5.8	5.0		
30	58.9	5.6		6.0	4.7		
31	59.3	5.5		6.2	4.5		
32	59.7	5.4		6.4	4.3		
33	60.1	5.4		6.6	4.0		
34	60.5	5.3		6.8	3.8		
35	60.9	5.2		7.0	3.5		
36	61.3	5.1		7.2	3.3		
37	61.7	5.0		7.4	3.0		
38	62.1	5.0		7.6	2.8		
39	62.5	4.9		7.8	2.6		
40	62.9	4.8		8.0	2.3		
41		4.7					
42		4.6					
43		4.6					
44		4.5					
45		4.4					
46		4.3					
47		4.2					
48		4.2					
49		4.1					
50		4.0					

nation of the test scores that the weighted values of Tests 5, 8, and 9 (the three with negative weights) would never exceed 8, 12, and 12, respectively. Accordingly, these three positive numbers have been subtracted from the constant term (78.9) and added to the corresponding negatively weighted scores, leaving of course in all cases a positive value. The remainder of the constant term (46.9) has been added to the weighted value of Test 1 and thus the constant term has been entirely eliminated as a separate value. The result of this splitting up of the constant term in the case of the forecast just considered is as follows :

$$\begin{array}{rcl}
 6.8 + 46.9 & = & 53.7 \\
 - 2.0 + 8.0 & = & 6.0 \\
 8.64 + 0.0 & = & 8.6 \\
 - 3.12 + 12.0 & = & 8.9 \\
 - 4.953 + 12.0 & = & 7.0 \\
 \hline
 & & 84.2
 \end{array}$$

Accordingly, in the facilitating table all the entries of weighted scores of Test 1 have 46.9 added, all those of Test 5 have 8 added, and all those of Tests 8 and 9 have 12 added. The use of the facilitating table thus eliminates not only all multiplication and all negative values, but also the constant term as an extra number to be added.

It should be noted that to make the table more convenient to use, entries for Test 8 have been made at intervals of .2 and for Test 9 at intervals of .05. If the tabular value nearest to any given test score be used, the result will be quite close enough for practical purposes.

It seems rather doubtful whether prediction formulæ as mathematical equations will attain much popularity among individuals not trained in the technical aspects of aptitude work. To the uninitiated, almost any formula, however simple, is provocative of a kind of terror. Yet once the test

batteries have been worked out by experts, individuals with relatively little technical training must be able to administer them if they are to be a practical success. It is believed that facilitating tables such as that described above, printed with contrasting type faces to make them easy to read, on cards convenient for handling, will make the combining of test scores by scientific methods almost as simple as combining them in the haphazard manner that is now so common.

A MACHINE WHICH MAKES APTITUDE FORECASTS AUTOMATICALLY

Another system of making aptitude predictions from forecasting formulæ has been devised by the writer (35) in the form of an automatic machine. This machine is an integral part of a comprehensive program of vocational guidance first sketched in 1923 (35 a). The program calls for the construction of a single universal battery of tests which shall sample, so far as possible, all of the important aptitude determiners. The battery will contain perhaps thirty or forty different test units and may require a day or more to administer. Upon the basis of this one battery there will be constructed separate forecasting formulæ for each of the more important type occupations — possibly to the number of forty or fifty. Thus there would be forty or fifty different equations, each equation weighting the tests of the one battery in a different way so as to make the best possible forecast of a particular aptitude. These equations would, of course, all be much longer than the one given above for free-hand drawing, each probably involving every one of the thirty or more tests of the battery. To make all the forty or fifty forecasts by such a system in the ordinary way would involve something like 1500 multiplications, all of which would need to be summated in a more or less complicated manner. This would represent

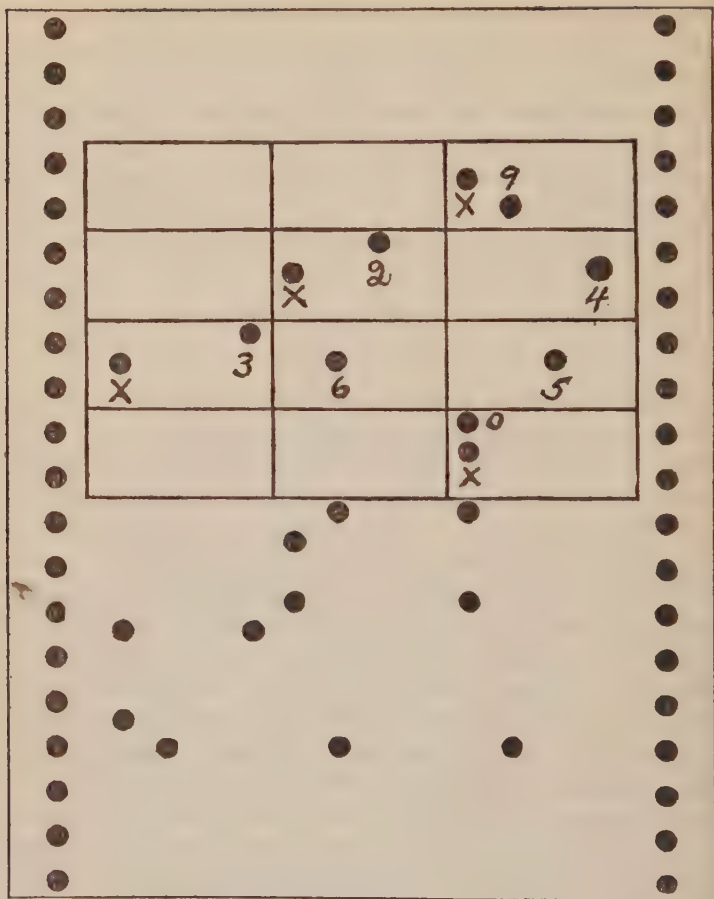


FIG. 60. A section of perforated paper tape (very slightly reduced) upon which test scores are recorded. The first four scores have been ruled off and the several digits indicated for purposes of explanation. The first score is 9, the second 24, the third 365, and so on. The lower portion shows the normal appearance of the tape. The perforations on the edges are to insure the accurate feeding of the tape through the aptitude-prediction machine.

a huge amount of labor, to say nothing of the human errors certain to creep into such a large amount of hand work. The forecasting machine mentioned above has been designed to perform this work automatically. Three of these machines have been constructed (34).

In its final form, this machine will have the different forecasting formulæ placed in it permanently as a four-inch perforated band of thin metal, somewhat resembling a music roll in appearance. The test scores will be given to the machine in the form of a similar perforated band of paper (Fig. 60) upon which have been recorded the test scores of a given subject. The test scores are recorded on this paper band by means of a special perforating device which is operated somewhat like a typewriter. A series of forty test scores may be thus recorded in about a minute. Once the test scores have been recorded on the paper band, it will be placed in the forecasting machine and the starter pressed. The machine will then proceed automatically, and without any attention whatever from the attendant, to make one aptitude forecast after another until the entire forty or fifty have been calculated.

At the time of inserting the band of test scores there will also be placed in another part of the machine a card bearing the name of the subject and a blank form giving in a column the names of all the aptitudes and occupations for which forecasting formulæ are available. As the machine makes its forecasts, it will stamp them down on this card automatically, opposite the names of the appropriate aptitudes. When the forecasts have all been made the machine will stop automatically, at the same time ringing a bell to call the attendant. The card of forecasts, when removed from the machine, will then present in orderly array and in units of a single uniform scale, permitting of instant comparison, forecasts of the individual's probable success in all the chief type occupations of the world. The youth whose potential aptitudes are thus

recorded may then examine the card to learn those vocations in which his chance of success is low. These may be avoided in his choice of a life work. He may then examine the card to learn those vocations in which his chance of success is greatest. The three or four most promising vocations thus emerging may be given further investigation. From these, in the light of his interests, opportunities, and general circumstances, may finally be chosen a life work.

It scarcely needs to be pointed out that the program of vocational guidance thus briefly sketched is a revolutionary departure from the current development of aptitude testing (35 a). This being the case, there will no doubt be considerable inertia and resistance from conservative quarters. To this difficulty must be added the fact that the program involves a vast amount of minutely coördinated research quite impossible of accomplishment by isolated workers. But the logic of the situation is certain to triumph in the end. We may look forward with confidence to a day not far distant when some such system as that sketched above will be operating in every large school system. Then, and not until then, will there be possible a genuine vocational guidance for the masses of the people.

APPENDIX ONE

TABLE FOR CONVERTING RANKS INTO LINEAR SCORES

(Prepared by Selmer C. Larson)

11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Rank
8.3	8.3	8.4	8.5	8.5	8.6	8.6	8.7	8.7	8.7	8.8	8.8	8.8	8.9	8.9	8.9	8.9	9.0	9.0	9.0	1
7.2	7.3	7.3	7.4	7.5	7.6	7.6	7.7	7.8	7.8	7.9	7.9	7.9	8.0	8.0	8.1	8.1	8.1	8.2	8.2	2
6.5	6.6	6.7	6.8	6.9	7.0	7.1	7.1	7.2	7.3	7.3	7.4	7.4	7.5	7.5	7.6	7.6	7.7	7.7	7.7	3
5.9	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	6.9	7.0	7.0	7.1	7.1	7.2	7.2	7.3	7.3	7.4	4
5.5	5.6	5.8	5.9	6.0	6.1	6.2	6.3	6.4	6.5	6.6	6.6	6.7	6.7	6.8	6.9	6.9	7.0	7.0	7.1	5
5.0	5.2	5.4	5.5	5.7	5.8	5.9	6.0	6.1	6.2	6.3	6.3	6.4	6.4	6.5	6.6	6.6	6.7	6.7	6.8	6
4.5	4.8	5.0	5.2	5.3	5.5	5.6	5.7	5.8	5.9	6.0	6.1	6.1	6.2	6.3	6.3	6.4	6.4	6.5	6.5	7
4.1	4.4	4.6	4.8	5.0	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0	6.0	6.1	6.2	6.2	6.3	6.3	8
3.5	3.9	4.2	4.5	4.7	4.8	5.0	5.1	5.3	5.4	5.5	5.6	5.7	5.7	5.8	5.9	6.0	6.0	6.1	6.1	9
2.8	3.4	3.8	4.1	4.3	4.5	4.7	4.9	5.0	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.8	5.9	5.9	10
1.7	2.7	3.3	3.7	4.0	4.2	4.4	4.6	4.7	4.9	5.0	5.1	5.2	5.3	5.4	5.5	5.6	5.6	5.7	5.8	11
	1.7	2.7	3.2	3.6	3.9	4.1	4.3	4.5	4.6	4.8	4.9	5.0	5.1	5.2	5.3	5.4	5.4	5.5	5.6	12
		1.6	2.6	3.1	3.5	3.8	4.0	4.2	4.4	4.5	4.7	4.8	4.9	5.0	5.1	5.2	5.3	5.3	5.4	13
			1.5	2.5	3.0	3.4	3.7	3.9	4.1	4.3	4.4	4.6	4.7	4.8	4.9	5.0	5.1	5.2	5.3	14
				1.5	2.4	2.9	3.3	3.6	3.8	4.0	4.2	4.3	4.5	4.6	4.7	4.8	4.9	5.0	5.1	15
					1.4	2.4	2.8	3.2	3.5	3.7	3.9	4.1	4.3	4.4	4.5	4.6	4.7	4.8	4.9	16
						1.4	2.3	2.8	3.1	3.4	3.7	3.9	4.0	4.2	4.3	4.4	4.6	4.7	4.7	17
							1.3	2.2	2.7	3.1	3.4	3.6	3.8	4.0	4.1	4.2	4.4	4.5	4.5	18
								1.3	2.2	2.7	3.0	3.3	3.6	3.7	3.9	4.0	4.2	4.3	4.4	19
									1.3	2.1	2.6	3.0	3.3	3.5	3.7	3.8	4.0	4.1	4.2	20
										1.2	2.1	2.6	2.9	3.2	3.4	3.6	3.8	3.9	4.1	21
											1.2	2.1	2.5	2.9	3.1	3.4	3.6	3.7	3.9	22
												1.2	2.0	2.5	2.8	3.1	3.4	3.5	3.7	23
													1.1	2.0	2.4	2.8	3.1	3.3	3.5	24
														1.1	1.9	2.4	2.7	3.0	3.3	25
															1.1	1.9	2.4	2.7	3.0	26
																1.1	1.9	2.3	2.7	27
																	1.0	1.8	2.3	28
																		1.0	1.8	29
																			1.0	30

How to Use the Table

This table converts the ranks of any series between 11 and 50 directly into units of amount on a 10-point scale. If the decimal points in the body of the tables are ignored, the conversion is into units of a 100-point scale. The number of individuals in the ranked group is found in the line of numbers in heavy-faced type at the top, and the rank of any individual within the group is found in the column of heavy-faced type at the right. To find the linear score of an individual ranked 23 in a group of 28, find column 28 at the top and follow down to row 23. The linear score thus found is 3.4.

TABLE FOR CONVERTING RANKS INTO LINEAR SCORES — *Continued*

31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	Rank
9.0	9.0	9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.2	9.2	9.2	9.2	9.2	9.2	9.2	9.3	9.3	9.3	9.3	1
8.2	8.2	8.3	8.3	8.3	8.3	8.4	8.4	8.4	8.4	8.5	8.5	8.5	8.5	8.5	8.5	8.6	8.6	8.6	8.6	2
7.7	7.8	7.8	7.8	7.9	7.9	7.9	7.9	8.0	8.0	8.0	8.0	8.1	8.1	8.1	8.1	8.1	8.2	8.2	8.2	3
7.4	7.4	7.5	7.5	7.6	7.6	7.6	7.6	7.7	7.7	7.7	7.7	7.7	7.7	7.8	7.8	7.8	7.8	7.8	7.9	4
7.1	7.1	7.2	7.2	7.3	7.3	7.3	7.3	7.3	7.4	7.4	7.4	7.5	7.5	7.5	7.5	7.6	7.6	7.6	7.6	5
6.8	6.9	6.9	6.9	7.0	7.0	7.0	7.1	7.1	7.1	7.2	7.2	7.2	7.3	7.3	7.3	7.3	7.4	7.4	7.4	6
6.6	6.6	6.7	6.7	6.8	6.8	6.8	6.9	6.9	6.9	7.0	7.0	7.0	7.1	7.1	7.1	7.1	7.2	7.2	7.2	7
6.4	6.4	6.5	6.5	6.6	6.6	6.6	6.7	6.7	6.7	6.8	6.8	6.8	6.9	6.9	6.9	7.0	7.0	7.0	7.0	8
6.2	6.3	6.3	6.4	6.4	6.4	6.5	6.5	6.5	6.6	6.6	6.6	6.7	6.7	6.7	6.8	6.8	6.8	6.8	6.9	9
6.0	6.1	6.1	6.2	6.2	6.2	6.3	6.3	6.4	6.4	6.4	6.5	6.5	6.5	6.6	6.6	6.6	6.7	6.7	6.7	10
5.8	5.9	5.9	6.0	6.0	6.1	6.1	6.2	6.2	6.3	6.3	6.4	6.4	6.4	6.5	6.5	6.5	6.5	6.6	6.6	11
5.7	5.7	5.8	5.8	5.9	5.9	6.0	6.0	6.1	6.1	6.1	6.2	6.2	6.2	6.2	6.2	6.2	6.3	6.3	6.4	12
5.5	5.5	5.6	5.7	5.7	5.8	5.8	5.9	5.9	6.0	6.0	6.0	6.1	6.1	6.2	6.2	6.2	6.3	6.3	6.4	13
5.3	5.4	5.5	5.5	5.6	5.6	5.7	5.7	5.8	5.8	5.9	5.9	6.0	6.0	6.0	6.1	6.1	6.1	6.2	6.2	14
5.2	5.2	5.3	5.4	5.4	5.5	5.5	5.6	5.6	5.7	5.7	5.8	5.8	5.9	5.9	5.9	6.0	6.0	6.1	6.1	15
5.0	5.1	5.2	5.2	5.3	5.3	5.4	5.5	5.5	5.6	5.6	5.7	5.7	5.8	5.8	5.9	5.9	5.9	6.0	6.0	16
4.8	4.9	5.0	5.1	5.1	5.2	5.3	5.3	5.4	5.4	5.5	5.5	5.6	5.6	5.7	5.7	5.8	5.8	5.8	5.9	17
4.7	4.8	4.8	4.9	5.0	5.1	5.1	5.2	5.3	5.3	5.4	5.4	5.5	5.5	5.6	5.6	5.6	5.7	5.7	5.8	18
4.5	4.6	4.7	4.8	4.9	4.9	5.0	5.1	5.1	5.2	5.2	5.3	5.3	5.4	5.4	5.5	5.5	5.5	5.6	5.6	19
4.4	4.5	4.5	4.6	4.7	4.8	4.9	4.9	5.0	5.1	5.1	5.2	5.2	5.3	5.3	5.4	5.4	5.4	5.5	5.6	20
4.2	4.3	4.4	4.5	4.6	4.7	4.7	4.8	4.9	4.9	5.0	5.1	5.1	5.2	5.2	5.3	5.3	5.4	5.4	5.4	21
4.0	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.7	4.8	4.9	4.9	5.0	5.1	5.1	5.2	5.2	5.3	5.3	5.3	22
3.8	3.9	4.1	4.2	4.3	4.4	4.4	4.5	4.6	4.7	4.8	4.8	4.9	4.9	5.0	5.1	5.1	5.2	5.2	5.2	23
3.6	3.8	3.9	4.0	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.7	4.8	4.8	4.9	4.9	5.0	5.0	5.1	5.1	24
3.4	3.6	3.7	3.8	4.0	4.1	4.2	4.3	4.4	4.4	4.5	4.6	4.7	4.7	4.8	4.8	4.9	4.9	5.0	5.0	25
3.2	3.4	3.5	3.6	3.8	3.9	4.0	4.1	4.2	4.3	4.4	4.5	4.5	4.6	4.7	4.7	4.8	4.8	4.9	5.0	26
2.9	3.2	3.3	3.4	3.6	3.8	3.9	4.0	4.1	4.2	4.3	4.4	4.4	4.5	4.6	4.6	4.7	4.7	4.8	4.9	27
2.6	2.9	3.1	3.2	3.4	3.6	3.7	3.8	3.9	4.0	4.1	4.2	4.3	4.4	4.4	4.5	4.6	4.6	4.7	4.8	28
2.3	2.6	2.9	3.0	3.2	3.4	3.5	3.7	3.8	3.9	4.0	4.1	4.2	4.3	4.3	4.4	4.5	4.5	4.6	4.7	29
1.8	2.2	2.6	2.8	3.0	3.2	3.3	3.5	3.6	3.8	3.9	4.0	4.0	4.1	4.2	4.3	4.4	4.4	4.5	4.6	30
1.0	1.8	2.2	2.5	2.8	3.0	3.1	3.3	3.5	3.6	3.7	3.8	3.9	4.0	4.1	4.2	4.3	4.4	4.4	4.5	31
	1.0	1.7	2.2	2.5	2.8	2.9	3.1	3.3	3.4	3.6	3.7	3.8	3.9	4.0	4.1	4.2	4.3	4.3	4.3	32
		.9	1.7	2.1	2.5	2.7	2.9	3.1	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0	4.1	4.2	4.2	33
			.9	1.7	2.1	2.4	2.7	2.9	3.1	3.2	3.4	3.5	3.6	3.7	3.8	3.9	4.0	4.1	4.1	34
				.9	1.7	2.1	2.4	2.7	2.9	3.0	3.2	3.3	3.5	3.6	3.7	3.8	3.9	3.9	4.0	35
					.9	1.6	2.1	2.4	2.6	2.8	3.0	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	36
						.9	1.6	2.0	2.3	2.6	2.8	3.0	3.1	3.3	3.4	3.5	3.6	3.7	3.8	37
							.9	1.6	2.0	2.3	2.6	2.8	2.9	3.1	3.2	3.4	3.5	3.6	3.7	38
								.8	1.6	2.0	2.3	2.5	2.7	2.9	3.1	3.2	3.3	3.4	3.5	39
									.8	1.6	2.0	2.3	2.5	2.7	2.9	3.0	3.2	3.3	3.4	40
										.8	1.5	1.9	2.3	2.5	2.7	2.9	3.0	3.1	3.3	41
											.8	1.5	1.9	2.2	2.5	2.7	2.8	3.0	3.1	42
												.8	1.5	1.9	2.2	2.4	2.6	2.8	3.0	43
													.8	1.5	1.9	2.2	2.4	2.6	2.8	44
														.8	1.5	1.9	2.2	2.4	2.6	45
															.8	1.4	1.9	2.2	2.4	46
																.7	1.4	1.8	2.1	47
																	.7	1.4	1.8	48
																		.7	1.4	49
																			.7	50

APPENDIX TWO

TABLE OF $1-r^2$ AND $\sqrt{1-r^2}$

r	$1 - r^2$	$\sqrt{1 - r^2}$	r	$1 - r^2$	$\sqrt{1 - r^2}$
.00	1.0000	1.0000	.25	.9375	.9682
.005	1.0000	1.0000	.255	.9350	.9670
.01	.9999	.9999	.26	.9324	.9656
.015	.9998	.9999	.265	.9298	.9643
.02	.9996	.9998	.27	.9271	.9629
.025	.9994	.9997	.275	.9244	.9615
.03	.9991	.9995	.28	.9216	.9600
.035	.9988	.9994	.285	.9188	.9585
.04	.9984	.9992	.29	.9159	.9570
.045	.9980	.9990	.295	.9130	.9555
.05	.9975	.9987	.30	.9100	.9539
.055	.9970	.9985	.305	.9070	.9524
.06	.9964	.9982	.31	.9039	.9507
.065	.9958	.9979	.315	.9008	.9491
.07	.9951	.9975	.32	.8976	.9474
.075	.9944	.9972	.325	.8944	.9457
.08	.9936	.9968	.33	.8911	.9440
.085	.9928	.9964	.335	.8878	.9422
.09	.9919	.9959	.34	.8844	.9404
.095	.9910	.9955	.345	.8810	.9386
.10	.9900	.9950	.35	.8775	.9367
.105	.9890	.9945	.355	.8740	.9349
.11	.9879	.9939	.36	.8704	.9330
.115	.9868	.9934	.365	.8668	.9310
.12	.9856	.9928	.37	.8631	.9290
.125	.9844	.9922	.375	.8594	.9270
.13	.9831	.9915	.38	.8556	.9250
.135	.9818	.9909	.385	.8518	.9229
.14	.9804	.9902	.39	.8479	.9208
.145	.9790	.9894	.395	.8440	.9187
.15	.9775	.9887	.40	.8400	.9165
.155	.9760	.9879	.405	.8360	.9143
.16	.9744	.9871	.41	.8319	.9121
.165	.9728	.9863	.415	.8278	.9098
.17	.9711	.9854	.42	.8236	.9075
.175	.9694	.9846	.425	.8194	.9052
.18	.9676	.9837	.43	.8151	.9028
.185	.9658	.9828	.435	.8108	.9004
.19	.9639	.9818	.44	.8064	.8980
.195	.9620	.9808	.445	.8020	.8955
.20	.9600	.9798	.45	.7975	.8930
.205	.9580	.9788	.455	.7930	.8905
.21	.9559	.9777	.46	.7884	.8879
.215	.9538	.9766	.465	.7838	.8853
.22	.9516	.9755	.47	.7791	.8827
.225	.9494	.9744	.475	.7744	.8800
.23	.9471	.9732	.48	.7696	.8773
.235	.9448	.9720	.485	.7648	.8745
.24	.9424	.9708	.49	.7599	.8717
.245	.9400	.9695	.495	.7550	.8689

r	$1 - r^2$	$\sqrt{1 - r^2}$	r	$1 - r^2$	$\sqrt{1 - r^2}$
.50	.7500	.8660	.75	.4375	.6614
.505	.7450	.8631	.755	.4300	.6557
.51	.7399	.8617	.76	.4224	.6499
.515	.7348	.8572	.765	.4148	.6440
.52	.7296	.8542	.77	.4071	.6380
.525	.7244	.8511	.775	.3994	.6320
.53	.7191	.8480	.78	.3916	.6258
.535	.7138	.8449	.785	.3838	.6195
.54	.7084	.8417	.79	.3759	.6131
.545	.7030	.8385	.795	.3680	.6066
.55	.6975	.8352	.80	.3600	.6000
.555	.6920	.8319	.805	.3520	.5933
.56	.6864	.8285	.81	.3439	.5864
.565	.6808	.8251	.815	.3358	.5795
.57	.6751	.8216	.82	.3276	.5724
.575	.6694	.8182	.825	.3194	.5652
.58	.6636	.8146	.83	.3111	.5578
.585	.6578	.8110	.835	.3028	.5503
.59	.6519	.8074	.84	.2944	.5426
.595	.6460	.8037	.845	.2860	.5348
.60	.6400	.8000	.85	.2775	.5268
.605	.6340	.7962	.855	.2690	.5187
.61	.6279	.7924	.86	.2604	.5103
.615	.6218	.7885	.865	.2518	.5018
.62	.6156	.7846	.87	.2431	.4931
.625	.6094	.7806	.875	.2344	.4841
.63	.6031	.7766	.88	.2256	.4750
.635	.5968	.7725	.885	.2168	.4656
.64	.5904	.7684	.89	.2079	.4560
.645	.5840	.7642	.895	.1990	.4461
.65	.5775	.7599	.90	.1900	.4359
.655	.5710	.7556	.905	.1810	.4254
.66	.5644	.7513	.91	.1719	.4146
.665	.5578	.7469	.915	.1628	.4035
.67	.5511	.7424	.92	.1536	.3919
.675	.5444	.7378	.925	.1444	.3800
.68	.5376	.7332	.93	.1351	.3676
.685	.5308	.7286	.935	.1258	.3547
.69	.5239	.7238	.94	.1164	.3412
.695	.5170	.7190	.945	.1070	.3271
.70	.5100	.7141	.95	.0975	.3122
.705	.5030	.7092	.955	.0880	.2966
.71	.4959	.7042	.96	.0784	.2800
.715	.4888	.6991	.965	.0688	.2623
.72	.4816	.6940	.97	.0591	.2431
.725	.4744	.6888	.975	.0494	.2223
.73	.4671	.6834	.98	.0396	.1990
.735	.4598	.6781	.985	.0298	.1726
.74	.4524	.6726	.99	.0199	.1411
.745	.4450	.6671	.995	.0100	.1000

APPENDIX THREE

TABLE OF SQUARES AND SQUARE ROOTS

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
1	1	1.000	51	26 01	7.141	101	1 02 01	10.050
2	4	1.414	52	27 04	7.211	102	1 04 04	10.100
3	9	1.732	53	28 09	7.280	103	1 06 09	10.149
4	16	2.000	54	29 16	7.348	104	1 08 16	10.198
5	25	2.236	55	30 25	7.416	105	1 10 25	10.247
6	36	2.449	56	31 36	7.483	106	1 12 36	10.296
7	49	2.646	57	32 49	7.550	107	1 14 49	10.344
8	64	2.828	58	33 64	7.616	108	1 16 64	10.392
9	81	3.000	59	34 81	7.681	109	1 18 81	10.440
10	1 00	3.162	60	36 00	7.746	110	1 21 00	10.488
11	1 21	3.317	61	37 21	7.810	111	1 23 21	10.536
12	1 44	3.464	62	38 44	7.874	112	1 25 44	10.583
13	1 69	3.606	63	39 69	7.937	113	1 27 69	10.630
14	1 96	3.742	64	40 96	8.000	114	1 29 96	10.677
15	2 25	3.873	65	42 25	8.062	115	1 32 25	10.724
16	2 56	4.000	66	43 56	8.124	116	1 34 56	10.770
17	2 89	4.123	67	44 89	8.185	117	1 36 89	10.817
18	3 24	4.243	68	46 24	8.246	118	1 39 24	10.863
19	3 61	4.359	69	47 61	8.307	119	1 41 61	10.909
20	4 00	4.472	70	49 00	8.367	120	1 44 00	10.954
21	4 41	4.583	71	50 41	8.426	121	1 46 41	11.000
22	4 84	4.690	72	51 84	8.484	122	1 48 84	11.045
23	5 29	4.796	73	53 29	8.544	123	1 51 29	11.091
24	5 76	4.899	74	54 76	8.602	124	1 53 76	11.136
25	6 25	5.000	75	56 25	8.660	125	1 56 25	11.180
26	6 76	5.099	76	57 76	8.718	126	1 58 76	11.225
27	7 29	5.196	77	59 29	8.775	127	1 61 29	11.269
28	7 84	5.292	78	60 84	8.832	128	1 63 84	11.314
29	8 41	5.385	79	62 41	8.888	129	1 66 41	11.358
30	9 00	5.477	80	64 00	8.944	130	1 69 00	11.402
31	9 61	5.568	81	65 61	9.000	131	1 71 61	11.446
32	10 24	5.657	82	67 24	9.055	132	1 74 24	11.489
33	10 89	5.745	83	68 89	9.110	133	1 76 89	11.533
34	11 56	5.831	84	70 56	9.165	134	1 79 56	11.576
35	12 25	5.916	85	72 25	9.220	135	1 82 25	11.619
36	12 96	6.000	86	73 96	9.274	136	1 84 96	11.662
37	13 69	6.083	87	75 69	9.327	137	1 87 69	11.705
38	14 44	6.164	88	77 44	9.381	138	1 90 44	11.747
39	15 21	6.245	89	79 21	9.434	139	1 93 21	11.790
40	16 00	6.325	90	81 00	9.487	140	1 96 00	11.832
41	16 81	6.403	91	82 81	9.539	141	1 98 81	11.874
42	17 64	6.481	92	84 64	9.592	142	2 01 64	11.916
43	18 49	6.557	93	86 49	9.644	143	2 04 49	11.958
44	19 36	6.633	94	88 36	9.695	144	2 07 36	12.000
45	20 25	6.708	95	90 25	9.747	145	2 10 25	12.042
46	21 16	6.782	96	92 16	9.798	146	2 13 16	12.083
47	22 09	6.856	97	94 09	9.849	147	2 16 09	12.124
48	23 04	6.928	98	96 04	9.899	148	2 19 04	12.166
49	24 01	7.000	99	98 01	9.950	149	2 22 01	12.207
50	25 00	7.071	100	1 00 00	10.000	150	2 25 00	12.247

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
151	2 28 01	12.288	201	4 04 01	14.177	251	6 30 01	15.843
152	2 31 04	12.329	202	4 08 04	14.213	252	6 35 04	15.875
153	2 34 09	12.369	203	4 12 09	14.248	253	6 40 09	15.906
154	2 37 16	12.410	204	4 16 16	14.283	254	6 45 16	15.937
155	2 40 25	12.450	205	4 20 25	14.318	255	6 50 25	15.969
156	2 43 36	12.490	206	4 24 36	14.353	256	6 55 36	16.000
157	2 46 49	12.530	207	4 28 49	14.387	257	6 60 49	16.031
158	2 49 64	12.570	208	4 32 64	14.422	258	6 65 64	16.062
159	2 52 81	12.610	209	4 36 81	14.457	259	6 70 81	16.093
160	2 56 00	12.649	210	4 41 00	14.491	260	6 76 00	16.125
161	2 59 21	12.689	211	4 45 21	14.526	261	6 81 21	16.155
162	2 62 44	12.728	212	4 49 44	14.560	262	6 86 44	16.186
163	2 65 69	12.767	213	4 53 69	14.595	263	6 91 69	16.217
164	2 68 96	12.806	214	4 57 96	14.629	264	6 96 96	16.248
165	2 72 25	12.845	215	4 62 25	14.663	265	7 02 25	16.279
166	2 75 56	12.884	216	4 66 56	14.697	266	7 07 56	16.310
167	2 78 89	12.923	217	4 70 89	14.731	267	7 12 89	16.340
168	2 82 24	12.961	218	4 75 24	14.765	268	7 18 24	16.371
169	2 85 61	13.000	219	4 79 61	14.799	269	7 23 61	16.401
170	2 89 00	13.038	220	4 84 00	14.832	270	7 29 00	16.432
171	2 92 41	13.077	221	4 88 41	14.866	271	7 34 41	16.462
172	2 95 84	13.115	222	4 92 84	14.900	272	7 39 84	16.492
173	2 99 29	13.153	223	4 97 29	14.933	273	7 45 29	16.523
174	3 02 76	13.191	224	5 01 76	14.967	274	7 50 76	16.553
175	3 06 25	13.229	225	5 06 25	15.000	275	7 56 25	16.583
176	3 09 76	13.266	226	5 10 76	15.033	276	7 61 76	16.613
177	3 13 29	13.304	227	5 15 29	15.067	277	7 67 29	16.643
178	3 16 84	13.342	228	5 19 84	15.100	278	7 72 84	16.673
179	3 20 41	13.379	229	5 24 41	15.133	279	7 78 41	16.703
180	3 24 00	13.416	230	5 29 00	15.166	280	7 84 00	16.733
181	3 27 61	13.454	231	5 33 61	15.199	281	7 89 61	16.763
182	3 31 24	13.491	232	5 38 24	15.232	282	7 95 24	16.793
183	3 34 89	13.528	233	5 42 89	15.264	283	8 00 89	16.823
184	3 38 56	13.565	234	5 47 56	15.297	284	8 06 56	16.852
185	3 42 25	13.601	235	5 52 25	15.330	285	8 12 25	16.882
186	3 45 96	13.638	236	5 56 96	15.362	286	8 17 96	16.912
187	3 49 69	13.675	237	5 61 69	15.395	287	8 23 69	16.941
188	3 53 44	13.711	238	5 66 44	15.427	288	8 29 44	16.971
189	3 57 21	13.748	239	5 71 21	15.460	289	8 35 21	17.000
190	3 61 00	13.784	240	5 76 00	15.492	290	8 41 00	17.029
191	3 64 81	13.820	241	5 80 81	15.524	291	8 46 81	17.059
192	3 68 64	13.856	242	5 85 64	15.556	292	8 52 64	17.088
193	3 72 49	13.892	243	5 90 49	15.588	293	8 58 49	17.117
194	3 76 36	13.928	244	5 95 36	15.620	294	8 64 36	17.146
195	3 80 25	13.964	245	6 00 25	15.652	295	8 70 25	17.176
196	3 84 16	14.000	246	6 05 16	15.684	296	8 76 16	17.205
197	3 88 09	14.036	247	6 10 09	15.716	297	8 82 09	17.234
198	3 92 04	14.071	248	6 15 04	15.748	298	8 88 04	17.263
199	3 96 01	14.107	249	6 20 01	15.780	299	8 94 01	17.292
200	4 00 00	14.142	250	6 25 00	15.811	300	9 00 00	17.321

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
301	9 06 01	17.349	351	12 32 01	18.735	401	16 08 01	20.025
302	9 12 04	17.378	352	12 39 04	18.762	402	16 16 04	20.050
303	9 18 09	17.407	353	12 46 09	18.788	403	16 24 09	20.075
304	9 24 16	17.436	354	12 53 16	18.815	404	16 32 16	20.100
305	9 30 25	17.464	355	12 60 25	18.841	405	16 40 25	20.125
306	9 36 36	17.493	356	12 67 36	18.868	406	16 48 36	20.149
307	9 42 49	17.521	357	12 74 49	18.894	407	16 56 49	20.174
308	9 48 64	17.550	358	12 81 64	18.921	408	16 64 64	20.199
309	9 54 81	17.578	359	12 88 81	18.947	409	16 72 81	20.224
310	9 61 00	17.607	360	12 96 00	18.974	410	16 81 00	20.248
311	9 67 21	17.635	361	13 03 21	19.000	411	16 89 21	20.273
312	9 73 44	17.664	362	13 10 44	19.026	412	16 97 44	20.298
313	9 79 69	17.692	363	13 17 69	19.053	413	17 05 69	20.322
314	9 85 96	17.720	364	13 24 96	19.079	414	17 13 96	20.347
315	9 92 25	17.748	365	13 32 25	19.105	415	17 22 25	20.372
316	9 98 56	17.776	366	13 39 56	19.131	416	17 30 56	20.396
317	10 04 89	17.804	367	13 46 89	19.157	417	17 38 89	20.421
318	10 11 24	17.833	368	13 54 24	19.183	418	17 47 24	20.445
319	10 17 61	17.861	369	13 61 61	19.209	419	17 55 61	20.469
320	10 24 00	17.889	370	13 69 00	19.235	420	17 64 00	20.494
321	10 30 41	17.916	371	13 76 41	19.261	421	17 72 41	20.518
322	10 36 84	17.944	372	13 83 84	19.287	422	17 80 84	20.543
323	10 43 29	17.972	373	13 91 29	19.313	423	17 89 29	20.567
324	10 49 76	18.000	374	13 98 76	19.339	424	17 97 76	20.591
325	10 56 25	18.028	375	14 06 25	19.365	425	18 06 25	20.616
326	10 62 76	18.055	376	14 13 76	19.391	426	18 14 76	20.640
327	10 69 29	18.083	377	14 21 29	19.416	427	18 23 29	20.664
328	10 75 84	18.111	378	14 28 84	19.442	428	18 31 84	20.688
329	10 82 41	18.138	379	14 36 41	19.468	429	18 40 41	20.712
330	10 89 00	18.166	380	14 44 00	19.494	430	18 49 00	20.736
331	10 95 61	18.193	381	14 51 61	19.519	431	18 57 61	20.761
332	11 02 24	18.221	382	14 59 24	19.545	432	18 66 24	20.785
333	11 08 89	18.248	383	14 66 89	19.570	433	18 74 89	20.809
334	11 15 56	18.276	384	14 74 56	19.596	434	18 83 56	20.833
335	11 22 25	18.303	385	14 82 25	19.621	435	18 92 25	20.857
336	11 28 96	18.330	386	14 89 96	19.647	436	19 00 96	20.881
337	11 35 69	18.358	387	14 97 69	19.672	437	19 09 69	20.905
338	11 42 44	18.385	388	15 05 44	19.698	438	19 18 44	20.928
339	11 49 21	18.412	389	15 13 21	19.723	439	19 27 21	20.952
340	11 56 00	18.439	390	15 21 00	19.748	440	19 36 00	20.976
341	11 62 81	18.466	391	15 28 81	19.774	441	19 44 81	21.000
342	11 69 64	18.493	392	15 36 64	19.799	442	19 53 64	21.024
343	11 76 49	18.520	393	15 44 49	19.824	443	19 62 49	21.048
344	11 83 36	18.547	394	15 52 36	19.849	444	19 71 36	21.071
345	11 90 25	18.574	395	15 60 25	19.875	445	19 80 25	21.095
346	11 97 16	18.601	396	15 68 16	19.900	446	19 89 16	21.119
347	12 04 09	18.628	397	15 76 09	19.925	447	19 98 09	21.142
348	12 11 04	18.655	398	15 84 04	19.950	448	20 07 04	21.166
349	12 18 01	18.682	399	15 92 01	19.975	449	20 16 01	21.190
350	12 25 00	18.708	400	16 00 00	20.000	450	20 25 00	21.213

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
451	20 34 01	21.237	501	25 10 01	22.383	551	30 36 01	23.473
452	20 43 04	21.260	502	25 20 04	22.405	552	30 47 04	23.495
453	20 52 09	21.284	503	25 30 09	22.428	553	30 58 09	23.516
454	20 61 16	21.307	504	25 40 16	22.450	554	30 69 16	23.537
455	20 70 25	21.331	505	25 50 25	22.472	555	30 80 25	23.558
456	20 79 36	21.354	506	25 60 36	22.494	556	30 91 36	23.580
457	20 88 49	21.378	507	25 70 49	22.517	557	31 02 49	23.601
458	20 97 64	21.401	508	25 80 64	22.539	558	31 13 64	23.622
459	21 06 81	21.424	509	25 90 81	22.561	559	31 24 81	23.643
460	21 16 00	21.448	510	26 01 00	22.583	560	31 36 00	23.664
461	21 25 21	21.471	511	26 11 21	22.605	561	31 47 21	23.685
462	21 34 44	21.494	512	26 21 44	22.627	562	31 58 44	23.707
463	21 43 69	21.517	513	26 31 69	22.650	563	31 69 69	23.728
464	21 52 96	21.541	514	26 41 96	22.672	564	31 80 96	23.749
465	21 62 25	21.564	515	26 52 25	22.694	565	31 92 25	23.770
466	21 71 56	21.587	516	26 62 56	22.716	566	32 03 56	23.791
467	21 80 89	21.610	517	26 72 89	22.738	567	32 14 89	23.812
468	21 90 24	21.633	518	26 83 24	22.760	568	32 26 24	23.833
469	21 99 61	21.656	519	26 93 61	22.782	569	32 37 61	23.854
470	22 09 00	21.679	520	27 04 00	22.804	570	32 49 00	23.875
471	22 18 41	21.703	521	27 14 41	22.825	571	32 60 41	23.896
472	22 27 84	21.726	522	27 24 84	22.847	572	32 71 84	23.917
473	22 37 29	21.749	523	27 35 29	22.869	573	32 83 29	23.937
474	22 46 76	21.772	524	27 45 76	22.891	574	32 94 76	23.958
475	22 56 25	21.794	525	27 56 25	22.913	575	33 06 25	23.979
476	22 65 76	21.817	526	27 66 76	22.935	576	33 17 76	24.000
477	22 75 29	21.840	527	27 77 29	22.956	577	33 29 29	24.021
478	22 84 84	21.863	528	27 87 84	22.978	578	33 40 84	24.042
479	22 94 41	21.886	529	27 98 41	23.000	579	33 52 41	24.062
480	23 04 00	21.909	530	28 09 00	23.022	580	33 64 00	24.083
481	23 13 61	21.932	531	28 19 61	23.043	581	33 75 61	24.104
482	23 23 24	21.954	532	28 30 24	23.065	582	33 87 24	24.125
483	23 32 89	21.977	533	28 40 89	23.087	583	33 98 89	24.145
484	23 42 56	22.000	534	28 51 56	23.108	584	34 10 56	24.166
485	23 52 25	22.023	535	28 62 25	23.130	585	34 22 25	24.187
486	23 61 96	22.045	536	28 72 96	23.152	586	34 33 96	24.207
487	23 71 69	22.068	537	28 83 69	23.173	587	34 45 69	24.228
488	23 81 44	22.091	538	28 94 44	23.195	588	34 57 44	24.249
489	23 91 21	22.113	539	29 05 21	23.216	589	34 69 21	24.269
490	24 01 00	22.136	540	29 16 00	23.238	590	34 81 00	24.290
491	24 10 81	22.159	541	29 26 81	23.259	591	34 92 81	24.310
492	24 20 64	22.181	542	29 37 64	23.281	592	35 04 64	24.331
493	24 30 49	22.204	543	29 48 49	23.302	593	35 16 49	24.352
494	24 40 36	22.226	544	29 59 36	23.324	594	35 28 36	24.372
495	24 50 25	22.249	545	29 70 25	23.345	595	35 40 25	24.393
496	24 60 16	22.271	546	29 81 16	23.367	596	35 52 16	24.413
497	24 70 09	22.293	547	29 92 09	23.388	597	35 64 09	24.434
498	24 80 04	22.316	548	30 03 04	23.409	598	35 76 04	24.454
499	24 90 01	22.338	549	30 14 01	23.431	599	35 88 01	24.474
500	25 00 00	22.361	550	30 25 00	23.452	600	36 00 00	24.495

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
601	36 12 01	24.515	651	42 38 01	25.515	701	49 14 01	26.476
602	36 24 04	24.536	652	42 51 04	25.534	702	49 28 04	26.495
603	36 36 09	24.556	653	42 64 09	25.554	703	49 42 09	26.514
604	36 48 16	24.576	654	42 77 16	25.573	704	49 56 16	26.533
605	36 60 25	24.597	655	42 90 25	25.593	705	49 70 25	26.552
606	36 72 36	24.617	656	43 03 36	25.612	706	49 84 36	26.571
607	36 84 49	24.637	657	43 16 49	25.632	707	49 98 49	26.589
608	36 96 64	24.658	658	43 29 64	25.652	708	50 12 64	26.608
609	37 08 81	24.678	659	43 42 81	25.671	709	50 26 81	26.627
610	37 21 00	24.698	660	43 56 00	25.690	710	50 41 00	26.646
611	37 33 21	24.718	661	43 69 21	25.710	711	50 55 21	26.665
612	37 45 44	24.739	662	43 82 44	25.729	712	50 69 44	26.683
613	37 57 69	24.759	663	43 95 69	25.749	713	50 83 69	26.702
614	37 69 96	24.779	664	44 08 96	25.768	714	50 97 96	26.721
615	37 82 25	24.799	665	44 22 25	25.788	715	51 12 25	26.739
616	37 94 56	24.819	666	44 35 56	25.807	716	51 26 56	26.758
617	38 06 89	24.839	667	44 48 89	25.826	717	51 40 89	26.777
618	38 19 24	24.860	668	44 62 24	25.846	718	51 55 24	26.796
619	38 31 61	24.880	669	44 75 61	25.865	719	51 69 61	26.814
620	38 44 00	24.900	670	44 89 00	25.884	720	51 84 00	26.833
621	38 56 41	24.920	671	45 02 41	25.904	721	51 98 41	26.851
622	38 68 84	24.940	672	45 15 84	25.923	722	52 12 84	26.870
623	38 81 29	24.960	673	45 29 29	25.942	723	52 27 29	26.889
624	38 93 76	24.980	674	45 42 76	25.962	724	52 41 76	26.907
625	39 06 25	25.000	675	45 56 25	25.981	725	52 56 25	26.926
626	39 18 76	25.020	676	45 69 76	26.000	726	52 70 76	26.944
627	39 31 29	25.040	677	45 83 29	26.019	727	52 85 29	26.963
628	39 43 84	25.060	678	45 96 84	26.038	728	52 99 84	26.981
629	39 56 41	25.080	679	46 10 41	26.058	729	53 14 41	27.000
630	39 69 00	25.100	680	46 24 00	26.077	730	53 29 00	27.019
631	39 81 61	25.120	681	46 37 61	26.096	731	53 43 61	27.037
632	39 94 24	25.140	682	46 51 24	26.115	732	53 58 24	27.055
633	40 06 89	25.159	683	46 64 89	26.134	733	53 72 89	27.074
634	40 19 56	25.179	684	46 78 56	26.153	734	53 87 56	27.092
635	40 32 25	25.199	685	46 92 25	26.173	735	54 02 25	27.111
636	40 44 96	25.219	686	47 05 96	26.192	736	54 16 96	27.129
637	40 57 69	25.239	687	47 19 69	26.211	737	54 31 69	27.148
638	40 70 44	25.259	688	47 33 44	26.230	738	54 46 44	27.166
639	40 83 21	25.278	689	47 47 21	26.249	739	54 61 21	27.185
640	40 96 00	25.298	690	47 61 00	26.268	740	54 76 00	27.203
641	41 08 81	25.318	691	47 74 81	26.287	741	54 90 81	27.221
642	41 21 64	25.338	692	47 88 64	26.306	742	55 05 64	27.240
643	41 34 49	25.357	693	48 02 49	26.325	743	55 20 49	27.258
644	41 47 36	25.377	694	48 16 36	26.344	744	55 35 36	27.276
645	41 60 25	25.397	695	48 30 25	26.363	745	55 50 25	27.295
646	41 73 16	25.417	696	48 44 16	26.382	746	55 65 16	27.313
647	41 86 09	25.436	697	48 58 09	26.401	747	55 80 09	27.331
648	41 99 04	25.456	698	48 72 04	26.420	748	55 95 04	27.350
649	42 12 01	25.475	699	48 86 01	26.439	749	56 10 01	27.368
650	42 25 00	25.495	700	49 00 00	26.458	750	56 25 00	27.386

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
751	56 40 01	27.404	801	64 16 01	28.302	851	72 42 01	29.172
752	56 55 04	27.423	802	64 32 04	28.320	852	72 59 04	29.189
753	56 70 09	27.441	803	64 48 09	28.337	853	72 76 09	29.206
754	56 85 16	27.459	804	64 64 16	28.355	854	72 93 16	29.223
755	57 00 25	27.477	805	64 80 25	28.373	855	73 10 25	29.240
756	57 15 36	27.495	806	64 96 36	28.390	856	73 27 36	29.257
757	57 30 49	27.514	807	65 12 49	28.408	857	73 44 49	29.275
758	57 45 64	27.532	808	65 28 64	28.425	858	73 61 64	29.292
759	57 60 81	27.550	809	65 44 81	28.443	859	73 78 81	29.309
760	57 76 00	27.568	810	65 61 00	28.460	860	73 96 00	29.326
761	57 91 21	27.586	811	65 77 21	28.478	861	74 13 21	29.343
762	58 06 44	27.604	812	65 93 44	28.496	862	74 30 44	29.360
763	58 21 69	27.622	813	66 09 69	28.513	863	74 47 69	29.377
764	58 36 96	27.641	814	66 25 96	28.531	864	74 64 96	29.394
765	58 52 25	27.659	815	66 42 25	28.548	865	74 82 25	29.411
766	58 67 56	27.677	816	66 58 56	28.566	866	74 99 56	29.428
767	58 82 89	27.695	817	66 74 89	28.583	867	75 16 89	29.445
768	58 98 24	27.713	818	66 91 24	28.601	868	75 34 24	29.462
769	59 13 61	27.731	819	67 07 61	28.618	869	75 51 61	29.479
770	59 29 00	27.749	820	67 24 00	28.636	870	75 69 00	29.496
771	59 44 41	27.767	821	67 40 41	28.653	871	75 86 41	29.513
772	59 59 84	27.785	822	67 56 84	28.671	872	76 03 84	29.530
773	59 75 29	27.803	823	67 73 29	28.688	873	76 21 29	29.547
774	59 90 76	27.821	824	67 89 76	28.705	874	76 38 76	29.563
775	60 06 25	27.839	825	68 06 25	28.723	875	76 56 25	29.580
776	60 21 76	27.857	826	68 22 76	28.740	876	76 73 76	29.597
777	60 37 29	27.875	827	68 39 29	28.758	877	76 91 29	29.614
778	60 52 84	27.893	828	68 55 84	28.775	878	77 08 84	29.631
779	60 68 41	27.911	829	68 72 41	28.792	879	77 26 41	29.648
780	60 84 00	27.928	830	68 89 00	28.810	880	77 44 00	29.665
781	60 99 61	27.946	831	69 05 61	28.827	881	77 61 61	29.682
782	61 15 24	27.964	832	69 22 24	28.844	882	77 79 24	29.698
783	61 30 89	27.982	833	69 38 89	28.862	883	77 96 89	29.715
784	61 46 56	28.000	834	69 55 56	28.879	884	78 14 56	29.732
785	61 62 25	28.018	835	69 72 25	28.896	885	78 32 25	29.749
786	61 77 96	28.036	836	69 88 96	28.914	886	78 49 96	29.766
787	61 93 69	28.054	837	70 05 69	28.931	887	78 67 69	29.783
788	62 09 44	28.071	838	70 22 44	28.948	888	78 85 44	29.799
789	62 25 21	28.089	839	70 39 21	28.965	889	79 03 21	29.816
790	62 41 00	28.107	840	70 56 00	28.983	890	79 21 00	29.833
791	62 56 81	28.125	841	70 72 81	29.000	891	79 38 81	29.850
792	62 72 64	28.142	842	70 89 64	29.017	892	79 56 64	29.866
793	62 88 49	28.160	843	71 06 49	29.034	893	79 74 49	29.883
794	63 04 36	28.178	844	71 23 36	29.052	894	79 92 36	29.900
795	63 20 25	28.196	845	71 40 25	29.069	895	80 10 25	29.917
796	63 36 16	28.213	846	71 57 16	29.086	896	80 28 16	29.933
797	63 52 09	28.231	847	71 74 09	29.103	897	80 46 09	29.950
798	63 68 04	28.249	848	71 91 04	29.120	898	80 64 04	29.967
799	63 84 01	28.267	849	72 08 01	29.138	899	80 82 01	29.983
800	64 00 00	28.284	850	72 25 00	29.155	900	81 00 00	30.000

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
901	81 18 01	30.017	951	90 44 01	30.838
902	81 36 04	30.033	952	90 63 04	30.854
903	81 54 09	30.050	953	90 82 09	30.871
904	81 72 16	30.067	954	91 01 16	30.887
905	81 90 25	30.083	955	91 20 25	30.903
906	82 08 36	30.100	956	91 39 36	30.919
907	82 26 49	30.116	957	91 58 49	30.935
908	82 44 64	30.133	958	91 77 64	30.952
909	82 62 81	30.150	959	91 96 81	30.968
910	82 81 00	30.166	960	92 16 00	30.984
911	82 99 21	30.183	961	92 35 21	31.000
912	83 17 44	30.199	962	92 54 44	31.016
913	83 35 69	30.216	963	92 73 69	31.032
914	83 53 96	30.232	964	92 92 96	31.048
915	83 72 25	30.249	965	93 12 25	31.064
916	83 90 56	30.265	966	93 31 56	31.081
917	84 08 89	30.282	967	93 50 89	31.097
918	84 27 24	30.299	968	93 70 24	31.113
919	84 45 61	30.315	969	93 89 61	31.129
920	84 64 00	30.332	970	94 09 00	31.145
921	84 82 41	30.348	971	94 28 41	31.161
922	85 00 84	30.364	972	94 47 84	31.177
923	85 19 29	30.381	973	94 67 29	31.193
924	85 37 76	30.397	974	94 86 76	31.209
925	85 56 25	30.414	975	95 06 25	31.225
926	85 74 76	30.430	976	95 25 76	31.241
927	85 93 29	30.447	977	95 45 29	31.257
928	86 11 84	30.463	978	95 64 84	31.273
929	86 30 41	30.480	979	95 84 41	31.289
930	86 49 00	30.496	980	96 04 00	31.305
931	86 67 61	30.512	981	96 23 61	31.321
932	86 86 24	30.529	982	96 43 24	31.337
933	87 04 89	30.545	983	96 62 89	31.353
934	87 23 56	30.561	984	96 82 56	31.369
935	87 42 25	30.578	985	97 02 25	31.385
936	87 60 96	30.594	986	97 21 96	31.401
937	87 79 69	30.610	987	97 41 69	31.417
938	87 98 44	30.627	988	97 61 44	31.432
939	88 17 21	30.643	989	97 81 21	31.448
940	88 36 00	30.659	990	98 01 00	31.464
941	88 54 81	30.676	991	98 20 81	31.480
942	88 73 64	30.692	992	98 40 64	31.496
943	88 92 49	30.708	993	98 60 49	31.512
944	89 11 36	30.725	994	98 80 36	31.528
945	89 30 25	30.741	995	99 00 25	31.544
946	89 49 16	30.757	996	99 20 16	31.559
947	89 68 09	30.773	997	99 40 09	31.575
948	89 87 04	30.790	998	99 60 04	31.591
949	90 06 01	30.806	999	99 80 01	31.607
950	90 25 00	30.822	1000	100 00 00	31.623

APPENDIX FOUR

MULTIPLICATION TABLE TO 100 X 100

	2	3	4	5	6	7	8	9	10	
1	2	3	4	5	6	7	8	9	10	1
2	4	6	8	10	12	14	16	18	20	2
3	6	9	12	15	18	21	24	27	30	3
4	8	12	16	20	24	28	32	36	40	4
5	10	15	20	25	30	35	40	45	50	5
6	12	18	24	30	36	42	48	54	60	6
7	14	21	28	35	42	49	56	63	70	7
8	16	24	32	40	48	56	64	72	80	8
9	18	27	36	45	54	63	72	81	90	9
10	20	30	40	50	60	70	80	90	100	10
11	22	33	44	55	66	77	88	99	110	11
12	24	36	48	60	72	84	96	108	120	12
13	26	39	52	65	78	91	104	117	130	13
14	28	42	56	70	84	98	112	126	140	14
15	30	45	60	75	90	105	120	135	150	15
16	32	48	64	80	96	112	128	144	160	16
17	34	51	68	85	102	119	136	153	170	17
18	36	54	72	90	108	126	144	162	180	18
19	38	57	76	95	114	133	152	171	190	19
20	40	60	80	100	120	140	160	180	200	20
21	42	63	84	105	126	147	168	189	210	21
22	44	66	88	110	132	154	176	198	220	22
23	46	69	92	115	138	161	184	207	230	23
24	48	72	96	120	144	168	192	216	240	24
25	50	75	100	125	150	175	200	225	250	25
26	52	78	104	130	156	182	208	234	260	26
27	54	81	108	135	162	189	216	243	270	27
28	56	84	112	140	168	196	224	252	280	28
29	58	87	116	145	174	203	232	261	290	29
30	60	90	120	150	180	210	240	270	300	30
31	62	93	124	155	186	217	248	279	310	31
32	64	96	128	160	192	224	256	288	320	32
33	66	99	132	165	198	231	264	297	330	33
34	68	102	136	170	204	238	272	306	340	34
35	70	105	140	175	210	245	280	315	350	35
36	72	108	144	180	216	252	288	324	360	36
37	74	111	148	185	222	259	296	333	370	37
38	76	114	152	190	228	266	304	342	380	38
39	78	117	156	195	234	273	312	351	390	39
40	80	120	160	200	240	280	320	360	400	40
41	82	123	164	205	246	287	328	369	410	41
42	84	126	168	210	252	294	336	378	420	42
43	86	129	172	215	258	301	344	387	430	43
44	88	132	176	220	264	308	352	396	440	44
45	90	135	180	225	270	315	360	405	450	45
46	92	138	184	230	276	322	368	414	460	46
47	94	141	188	235	282	329	376	423	470	47
48	96	144	192	240	288	336	384	432	480	48
49	98	147	196	245	294	343	392	441	490	49
50	100	150	200	250	300	350	400	450	500	50
	2	3	4	5	6	7	8	9	10	

	2	3	4	5	6	7	8	9	10	
51	102	153	204	255	306	357	408	459	510	51
52	104	156	208	260	312	364	416	468	520	52
53	106	159	212	265	318	371	424	477	530	53
54	108	162	216	270	324	378	432	486	540	54
55	110	165	220	275	330	385	440	495	550	55
56	112	168	224	280	336	392	448	504	560	56
57	114	171	228	285	342	399	456	513	570	57
58	116	174	232	290	348	406	464	522	580	58
59	118	177	236	295	354	413	472	531	590	59
60	120	180	240	300	360	420	480	540	600	60
61	122	183	244	305	366	427	488	549	610	61
62	124	186	248	310	372	434	496	558	620	62
63	126	189	252	315	378	441	504	567	630	63
64	128	192	256	320	384	448	512	576	640	64
65	130	195	260	325	390	455	520	585	650	65
66	132	198	264	330	396	462	528	594	660	66
67	134	201	268	335	402	469	536	603	670	67
68	136	204	272	340	408	476	544	612	680	68
69	138	207	276	345	414	483	552	621	690	69
70	140	210	280	350	420	490	560	630	700	70
71	142	213	284	355	426	497	568	639	710	71
72	144	216	288	360	432	504	576	648	720	72
73	146	219	292	365	438	511	584	657	730	73
74	148	222	296	370	444	518	592	666	740	74
75	150	225	300	375	450	525	600	675	750	75
76	152	228	304	380	456	532	608	684	760	76
77	154	231	308	385	462	539	616	693	770	77
78	156	234	312	390	468	546	624	702	780	78
79	158	237	316	395	474	553	632	711	790	79
80	160	240	320	400	480	560	640	720	800	80
81	162	243	324	405	486	567	648	729	810	81
82	164	246	328	410	492	574	656	738	820	82
83	166	249	332	415	498	581	664	747	830	83
84	168	252	336	420	504	588	672	756	840	84
85	170	255	340	425	510	595	680	765	850	85
86	172	258	344	430	516	602	688	774	860	86
87	174	261	348	435	522	609	696	783	870	87
88	176	264	352	440	528	616	704	792	880	88
89	178	267	356	445	534	623	712	801	890	89
90	180	270	360	450	540	630	720	810	900	90
91	182	273	364	455	546	637	728	819	910	91
92	184	276	368	460	552	644	736	828	920	92
93	186	279	372	465	558	651	744	837	930	93
94	188	282	376	470	564	658	752	846	940	94
95	190	285	380	475	570	665	760	855	950	95
96	192	288	384	480	576	672	768	864	960	96
97	194	291	388	485	582	679	776	873	970	97
98	196	294	392	490	588	686	784	882	980	98
99	198	297	396	495	594	693	792	891	990	99
100	200	300	400	500	600	700	800	900	1000	100
	2	3	4	5	6	7	8	9	10	

	11	12	13	14	15	16	17	18	19	20	
1	11	12	13	14	15	16	17	18	19	20	1
2	22	24	26	28	30	32	34	36	38	40	2
3	33	36	39	42	45	48	51	54	57	60	3
4	44	48	52	56	60	64	68	72	76	80	4
5	55	60	65	70	75	80	85	90	95	100	5
6	66	72	78	84	90	96	102	108	114	120	6
7	77	84	91	98	105	112	119	126	133	140	7
8	88	96	104	112	120	128	136	144	152	160	8
9	99	108	117	126	135	144	153	162	171	180	9
10	110	120	130	140	150	160	170	180	190	200	10
11	121	132	143	154	165	176	187	198	209	220	11
12	132	144	156	168	180	192	204	216	228	240	12
13	143	156	169	182	195	208	221	234	247	260	13
14	154	168	182	196	210	224	238	252	266	280	14
15	165	180	195	210	225	240	255	270	285	300	15
16	176	192	208	224	240	256	272	288	304	320	16
17	187	204	221	238	255	272	289	306	323	340	17
18	198	216	234	252	270	288	306	324	342	360	18
19	209	228	247	266	285	304	323	342	361	380	19
20	220	240	260	280	300	320	340	360	380	400	20
21	231	252	273	294	315	336	357	378	399	420	21
22	242	264	286	308	330	352	374	396	418	440	22
23	253	276	299	322	345	368	391	414	437	460	23
24	264	288	312	336	360	384	408	432	456	480	24
25	275	300	325	350	375	400	425	450	475	500	25
26	286	312	338	364	390	416	442	468	494	520	26
27	297	324	351	378	405	432	459	486	513	540	27
28	308	336	364	392	420	448	476	504	532	560	28
29	319	348	377	406	435	464	493	522	551	580	29
30	330	360	390	420	450	480	510	540	570	600	30
31	341	372	403	434	465	496	527	558	589	620	31
32	352	384	416	448	480	512	544	576	608	640	32
33	363	396	429	462	495	528	561	594	627	660	33
34	374	408	442	476	510	544	578	612	646	680	34
35	385	420	455	490	525	560	595	630	665	700	35
36	396	432	468	504	540	576	612	648	684	720	36
37	407	444	481	518	555	592	629	666	703	740	37
38	418	456	494	532	570	608	646	684	722	760	38
39	429	468	507	546	585	624	663	702	741	780	39
40	440	480	520	560	600	640	680	720	760	800	40
41	451	492	533	574	615	656	697	738	779	820	41
42	462	504	546	588	630	672	714	756	798	840	42
43	473	516	559	602	645	688	731	774	817	860	43
44	484	528	572	616	660	704	748	792	836	880	44
45	495	540	585	630	675	720	765	810	855	900	45
46	506	552	598	644	690	736	782	828	874	920	46
47	517	564	611	658	705	752	799	846	893	940	47
48	528	576	624	672	720	768	816	864	912	960	48
49	539	588	637	686	735	784	833	882	931	980	49
50	550	600	650	700	750	800	850	900	950	1000	50
	11	12	13	14	15	16	17	18	19	20	

Multiplication Table

505

	11	12	13	14	15	16	17	18	19	20	
51	561	612	663	714	765	816	867	918	969	1020	51
52	572	624	676	728	780	832	884	936	988	1040	52
53	583	636	689	742	795	848	901	954	1007	1060	53
54	594	648	702	756	810	864	918	972	1026	1080	54
55	605	660	715	770	825	880	935	990	1045	1100	55
56	616	672	728	784	840	896	952	1008	1064	1120	56
57	627	684	741	798	855	912	969	1026	1083	1140	57
58	638	696	754	812	870	928	986	1044	1102	1160	58
59	649	708	767	826	885	944	1003	1062	1121	1180	59
60	660	720	780	840	900	960	1020	1080	1140	1200	60
61	671	732	793	854	915	976	1037	1098	1159	1220	61
62	682	744	806	868	930	992	1054	1116	1178	1240	62
63	693	756	819	882	945	1008	1071	1134	1197	1260	63
64	704	768	832	896	960	1024	1088	1152	1216	1280	64
65	715	780	845	910	975	1040	1105	1170	1235	1300	65
66	726	792	858	924	990	1056	1122	1188	1254	1320	66
67	737	804	871	938	1005	1072	1139	1206	1273	1340	67
68	748	816	884	952	1020	1088	1156	1224	1292	1360	68
69	759	828	897	966	1035	1104	1173	1242	1311	1380	69
70	770	840	910	980	1050	1120	1190	1260	1330	1400	70
71	781	852	923	994	1065	1136	1207	1278	1349	1420	71
72	792	864	936	1008	1080	1152	1224	1296	1368	1440	72
73	803	876	949	1022	1095	1168	1241	1314	1387	1460	73
74	814	888	962	1036	1110	1184	1258	1332	1406	1480	74
75	825	900	975	1050	1125	1200	1275	1350	1425	1500	75
76	836	912	988	1064	1140	1216	1292	1368	1444	1520	76
77	847	924	1001	1078	1155	1232	1309	1386	1463	1540	77
78	858	936	1014	1092	1170	1248	1326	1404	1482	1560	78
79	869	948	1027	1106	1185	1264	1343	1422	1501	1580	79
80	880	960	1040	1120	1200	1280	1360	1440	1520	1600	80
81	891	972	1053	1134	1215	1296	1377	1458	1539	1620	81
82	902	984	1066	1148	1230	1312	1394	1476	1558	1640	82
83	913	996	1079	1162	1245	1328	1411	1494	1577	1660	83
84	924	1008	1092	1176	1260	1344	1428	1512	1596	1680	84
85	935	1020	1105	1190	1275	1360	1445	1530	1615	1700	85
86	946	1032	1118	1204	1290	1376	1462	1548	1634	1720	86
87	957	1044	1131	1218	1305	1392	1479	1566	1653	1740	87
88	968	1056	1144	1232	1320	1408	1496	1584	1672	1760	88
89	979	1068	1157	1246	1335	1424	1513	1602	1691	1780	89
90	990	1080	1170	1260	1350	1440	1530	1620	1710	1800	90
91	1001	1092	1183	1274	1365	1456	1547	1638	1729	1820	91
92	1012	1104	1196	1288	1380	1472	1564	1656	1748	1840	92
93	1023	1116	1209	1302	1395	1488	1581	1674	1767	1860	93
94	1034	1128	1222	1316	1410	1504	1598	1692	1786	1880	94
95	1045	1140	1235	1330	1425	1520	1615	1710	1805	1900	95
96	1056	1152	1248	1344	1440	1536	1632	1728	1824	1920	96
97	1067	1164	1261	1358	1455	1552	1649	1746	1843	1940	97
98	1078	1176	1274	1372	1470	1568	1666	1764	1862	1960	98
99	1089	1188	1287	1386	1485	1584	1683	1782	1881	1980	99
100	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	100
	11	12	13	14	15	16	17	18	19	20	

	21	22	23	24	25	26	27	28	29	30	
1	21	22	23	24	25	26	27	28	29	30	1
2	42	44	46	48	50	52	54	56	58	60	2
3	63	66	69	72	75	78	81	84	87	90	3
4	84	88	92	96	100	104	108	112	116	120	4
5	105	110	115	120	125	130	135	140	145	150	5
6	126	132	138	144	150	156	162	168	174	180	6
7	147	154	161	168	175	182	189	196	203	210	7
8	168	176	184	192	200	208	216	224	232	240	8
9	189	198	207	216	225	234	243	252	261	270	9
10	210	220	230	240	250	260	270	280	290	300	10
11	231	242	253	264	275	286	297	308	319	330	11
12	252	264	276	288	300	312	324	336	348	360	12
13	273	286	299	312	325	338	351	364	377	390	13
14	294	308	322	336	350	364	378	392	406	420	14
15	315	330	345	360	375	390	405	420	435	450	15
16	336	352	368	384	400	416	432	448	464	480	16
17	357	374	391	408	425	442	459	476	493	510	17
18	378	396	414	432	450	468	486	504	522	540	18
19	399	418	437	456	475	494	513	532	551	570	19
20	420	440	460	480	500	520	540	560	580	600	20
21	441	462	483	504	525	546	567	588	609	630	21
22	462	484	506	528	550	572	594	616	638	660	22
23	483	506	529	552	575	598	621	644	667	690	23
24	504	528	552	576	600	624	648	672	696	720	24
25	525	550	575	600	625	650	675	700	725	750	25
26	546	572	598	624	650	676	702	728	754	780	26
27	567	594	621	648	675	702	729	756	783	810	27
28	588	616	644	672	700	728	756	784	812	840	28
29	609	638	667	696	725	754	783	812	841	870	29
30	630	660	690	720	750	780	810	840	870	900	30
31	651	682	713	744	775	806	837	868	899	930	31
32	672	704	736	768	800	832	864	896	928	960	32
33	693	726	759	792	825	858	891	924	957	990	33
34	714	748	782	816	850	884	918	952	986	1020	34
35	735	770	805	840	875	910	945	980	1015	1050	35
36	756	792	828	864	900	936	972	1008	1044	1080	36
37	777	814	851	888	925	962	999	1036	1073	1110	37
38	798	836	874	912	950	988	1026	1064	1102	1140	38
39	819	858	897	936	975	1014	1053	1092	1131	1170	39
40	840	880	920	960	1000	1040	1080	1120	1160	1200	40
41	861	902	943	984	1025	1066	1107	1148	1189	1230	41
42	882	924	966	1008	1050	1092	1134	1176	1218	1260	42
43	903	946	989	1032	1075	1118	1161	1204	1247	1290	43
44	924	968	1012	1056	1100	1144	1188	1232	1276	1320	44
45	945	990	1035	1080	1125	1170	1215	1260	1305	1350	45
46	966	1012	1058	1104	1150	1196	1242	1288	1334	1380	46
47	987	1034	1081	1128	1175	1222	1269	1316	1363	1410	47
48	1008	1056	1104	1152	1200	1248	1296	1344	1392	1440	48
49	1029	1078	1127	1176	1225	1274	1323	1372	1421	1470	49
50	1050	1100	1150	1200	1250	1300	1350	1400	1450	1500	50
	21	22	23	24	25	26	27	28	29	30	

Multiplication Table

507

	21	22	23	24	25	26	27	28	29	30	
51	1071	1122	1173	1224	1275	1326	1377	1428	1479	1530	51
52	1092	1144	1196	1248	1300	1352	1404	1456	1508	1560	52
53	1113	1166	1219	1272	1325	1378	1431	1484	1537	1590	53
54	1134	1188	1242	1296	1350	1404	1458	1512	1566	1620	54
55	1155	1210	1265	1320	1375	1430	1485	1540	1595	1650	55
56	1176	1232	1288	1344	1400	1456	1512	1568	1624	1680	56
57	1197	1254	1311	1368	1425	1482	1539	1596	1653	1710	57
58	1218	1276	1334	1392	1450	1508	1566	1624	1682	1740	58
59	1239	1298	1357	1416	1475	1534	1593	1652	1711	1770	59
60	1260	1320	1380	1440	1500	1560	1620	1680	1740	1800	60
61	1281	1342	1403	1464	1525	1586	1647	1708	1769	1830	61
62	1302	1364	1426	1488	1550	1612	1674	1736	1798	1860	62
63	1323	1386	1449	1512	1575	1638	1701	1764	1827	1890	63
64	1344	1408	1472	1536	1600	1664	1728	1792	1856	1920	64
65	1365	1430	1495	1560	1625	1690	1755	1820	1885	1950	65
66	1386	1452	1518	1584	1650	1716	1782	1848	1914	1980	66
67	1407	1474	1541	1608	1675	1742	1809	1876	1943	2010	67
68	1428	1496	1564	1632	1700	1768	1836	1904	1972	2040	68
69	1449	1518	1587	1656	1725	1794	1863	1932	2001	2070	69
70	1470	1540	1610	1680	1750	1820	1890	1960	2030	2100	70
71	1491	1562	1633	1704	1775	1846	1917	1988	2059	2130	71
72	1512	1584	1656	1728	1800	1872	1944	2016	2088	2160	72
73	1533	1606	1679	1752	1825	1898	1971	2044	2117	2190	73
74	1554	1628	1702	1776	1850	1924	1998	2072	2146	2220	74
75	1575	1650	1725	1800	1875	1950	2025	2100	2175	2250	75
76	1596	1672	1748	1824	1900	1976	2052	2128	2204	2280	76
77	1617	1694	1771	1848	1925	2002	2079	2156	2233	2310	77
78	1638	1716	1794	1872	1950	2028	2106	2184	2262	2340	78
79	1659	1738	1817	1896	1975	2054	2133	2212	2291	2370	79
80	1680	1760	1840	1920	2000	2080	2160	2240	2320	2400	80
81	1701	1782	1863	1944	2025	2106	2187	2268	2349	2430	81
82	1722	1804	1886	1968	2050	2132	2214	2296	2378	2460	82
83	1743	1826	1909	1992	2075	2158	2241	2324	2407	2490	83
84	1764	1848	1932	2016	2100	2184	2268	2352	2436	2520	84
85	1785	1870	1955	2040	2125	2210	2295	2380	2465	2550	85
86	1806	1892	1978	2064	2150	2236	2322	2408	2494	2580	86
87	1827	1914	2001	2088	2175	2262	2349	2436	2523	2610	87
88	1848	1936	2024	2112	2200	2288	2376	2464	2552	2640	88
89	1869	1958	2047	2136	2225	2314	2403	2492	2581	2670	89
90	1890	1980	2070	2160	2250	2340	2430	2520	2610	2700	90
91	1911	2002	2093	2184	2275	2366	2457	2548	2639	2730	91
92	1932	2024	2116	2208	2300	2392	2484	2576	2668	2760	92
93	1953	2046	2139	2232	2325	2418	2511	2604	2697	2790	93
94	1974	2068	2162	2256	2350	2444	2538	2632	2726	2820	94
95	1995	2090	2185	2280	2375	2470	2565	2660	2755	2850	95
96	2016	2112	2208	2304	2400	2496	2592	2688	2784	2880	96
97	2037	2134	2231	2328	2425	2522	2619	2716	2813	2910	97
98	2058	2156	2254	2352	2450	2548	2646	2744	2842	2940	98
99	2079	2178	2277	2376	2475	2574	2673	2772	2871	2970	99
100	2100	2200	2300	2400	2500	2600	2700	2800	2900	3000	100
	21	22	23	24	25	26	27	28	29	30	

	31	32	33	34	35	36	37	38	39	40	
1	31	32	33	34	35	36	37	38	39	40	1
2	62	64	66	68	70	72	74	76	78	80	2
3	93	96	99	102	105	108	111	114	117	120	3
4	124	128	132	136	140	144	148	152	156	160	4
5	155	160	165	170	175	180	185	190	195	200	5
6	186	192	198	204	210	216	222	228	234	240	6
7	217	224	231	238	245	252	259	266	273	280	7
8	248	256	264	272	280	288	296	304	312	320	8
9	279	288	297	306	315	324	333	342	351	360	9
10	310	320	330	340	350	360	370	380	390	400	10
11	341	352	363	374	385	396	407	418	429	440	11
12	372	384	396	408	420	432	444	456	468	480	12
13	403	416	429	442	455	468	481	494	507	520	13
14	434	448	462	476	490	504	518	532	546	560	14
15	465	480	495	510	525	540	555	570	585	600	15
16	496	512	528	544	560	576	592	608	624	640	16
17	527	544	561	578	595	612	629	646	663	680	17
18	558	576	594	612	630	648	666	684	702	720	18
19	589	608	627	646	665	684	703	722	741	760	19
20	620	640	660	680	700	720	740	760	780	800	20
21	651	672	693	714	735	756	777	798	819	840	21
22	682	704	726	748	770	792	814	836	858	880	22
23	713	736	759	782	805	828	851	874	897	920	23
24	744	768	792	816	840	864	888	912	936	960	24
25	775	800	825	850	875	900	925	950	975	1000	25
26	806	832	858	884	910	936	962	988	1014	1040	26
27	837	864	891	918	945	972	999	1026	1053	1080	27
28	868	896	924	952	980	1008	1036	1064	1092	1120	28
29	899	928	957	986	1015	1044	1073	1102	1131	1160	29
30	930	960	990	1020	1050	1080	1110	1140	1170	1200	30
31	961	992	1023	1054	1085	1116	1147	1178	1209	1240	31
32	992	1024	1056	1088	1120	1152	1184	1216	1248	1280	32
33	1023	1056	1089	1122	1155	1188	1221	1254	1287	1320	33
34	1054	1088	1122	1156	1190	1224	1258	1292	1326	1360	34
35	1085	1120	1155	1190	1225	1260	1295	1330	1365	1400	35
36	1116	1152	1188	1224	1260	1296	1332	1368	1404	1440	36
37	1147	1184	1221	1258	1295	1332	1369	1406	1443	1480	37
38	1178	1216	1254	1292	1330	1368	1406	1444	1482	1520	38
39	1209	1248	1287	1326	1365	1404	1443	1482	1521	1560	39
40	1240	1280	1320	1360	1400	1440	1480	1520	1560	1600	40
41	1271	1312	1353	1394	1435	1476	1517	1558	1599	1640	41
42	1302	1344	1386	1428	1470	1512	1554	1596	1638	1680	42
43	1333	1376	1419	1462	1505	1548	1591	1634	1677	1720	43
44	1364	1408	1452	1496	1540	1584	1628	1672	1716	1760	44
45	1395	1440	1485	1530	1575	1620	1665	1710	1755	1800	45
46	1426	1472	1518	1564	1610	1656	1702	1748	1794	1840	46
47	1457	1504	1551	1598	1645	1692	1739	1786	1833	1880	47
48	1488	1536	1584	1632	1680	1728	1776	1824	1872	1920	48
49	1519	1568	1617	1666	1715	1764	1813	1862	1911	1960	49
50	1550	1600	1650	1700	1750	1800	1850	1900	1950	2000	50
	31	32	33	34	35	36	37	38	39	40	

	31	32	33	34	35	36	37	38	39	40	
51	1581	1632	1683	1734	1785	1836	1887	1938	1989	2040	51
52	1612	1664	1716	1768	1820	1872	1924	1976	2028	2080	52
53	1643	1696	1749	1802	1855	1908	1961	2014	2067	2120	53
54	1674	1728	1782	1836	1890	1944	1998	2052	2106	2160	54
55	1705	1760	1815	1870	1925	1980	2035	2090	2145	2200	55
56	1736	1792	1848	1904	1960	2016	2072	2128	2184	2240	56
57	1767	1824	1881	1938	1995	2052	2109	2166	2223	2280	57
58	1798	1856	1914	1972	2030	2088	2146	2204	2262	2320	58
59	1829	1888	1947	2006	2065	2124	2183	2242	2301	2360	59
60	1860	1920	1980	2040	2100	2160	2220	2280	2340	2400	60
61	1891	1952	2013	2074	2135	2196	2257	2318	2379	2440	61
62	1922	1984	2046	2108	2170	2232	2294	2356	2418	2480	62
63	1953	2016	2079	2142	2205	2268	2331	2394	2457	2520	63
64	1984	2048	2112	2176	2240	2304	2368	2432	2496	2560	64
65	2015	2080	2145	2210	2275	2340	2405	2470	2535	2600	65
66	2046	2112	2178	2244	2310	2376	2442	2508	2574	2640	66
67	2077	2144	2211	2278	2345	2412	2479	2546	2613	2680	67
68	2108	2176	2244	2312	2380	2448	2516	2584	2652	2720	68
69	2139	2208	2277	2346	2415	2484	2553	2622	2691	2760	69
70	2170	2240	2310	2380	2450	2520	2590	2660	2730	2800	70
71	2201	2272	2343	2414	2485	2556	2627	2698	2769	2840	71
72	2232	2304	2376	2448	2520	2592	2664	2736	2808	2880	72
73	2263	2336	2409	2482	2555	2628	2701	2774	2847	2920	73
74	2294	2368	2442	2516	2590	2664	2738	2812	2886	2960	74
75	2325	2400	2475	2550	2625	2700	2775	2850	2925	3000	75
76	2356	2432	2508	2584	2660	2736	2812	2888	2964	3040	76
77	2387	2464	2541	2618	2695	2772	2849	2926	3003	3080	77
78	2418	2496	2574	2652	2730	2808	2886	2964	3042	3120	78
79	2449	2528	2607	2686	2765	2844	2923	3002	3081	3160	79
80	2480	2560	2640	2720	2800	2880	2960	3040	3120	3200	80
81	2511	2592	2673	2754	2835	2916	2997	3078	3159	3240	81
82	2542	2624	2706	2788	2870	2952	3034	3116	3198	3280	82
83	2573	2656	2739	2822	2905	2988	3071	3154	3237	3320	83
84	2604	2688	2772	2856	2940	3024	3108	3192	3276	3360	84
85	2635	2720	2805	2890	2975	3060	3145	3230	3315	3400	85
86	2666	2752	2838	2924	3010	3096	3182	3268	3354	3440	86
87	2697	2784	2871	2958	3045	3132	3219	3306	3393	3480	87
88	2728	2816	2904	2992	3080	3168	3256	3344	3432	3520	88
89	2759	2848	2937	3026	3115	3204	3293	3382	3471	3560	89
90	2790	2880	2970	3060	3150	3240	3330	3420	3510	3600	90
91	2821	2912	3003	3094	3185	3276	3367	3458	3549	3640	91
92	2852	2944	3036	3128	3220	3312	3404	3496	3588	3680	92
93	2883	2976	3069	3162	3255	3348	3441	3534	3627	3720	93
94	2914	3008	3102	3196	3290	3384	3478	3572	3666	3760	94
95	2945	3040	3135	3230	3325	3420	3515	3610	3705	3800	95
96	2976	3072	3168	3264	3360	3456	3552	3648	3744	3840	96
97	3007	3104	3201	3298	3395	3492	3589	3686	3783	3880	97
98	3038	3136	3234	3332	3430	3528	3626	3724	3822	3920	98
99	3069	3168	3267	3366	3465	3564	3663	3762	3861	3960	99
100	3100	3200	3300	3400	3500	3600	3700	3800	3900	4000	100
	31	32	33	34	35	36	37	38	39	40	

	41	42	43	44	45	46	47	48	49	50	
1	41	42	43	44	45	46	47	48	49	50	1
2	82	84	86	88	90	92	94	96	98	100	2
3	123	126	129	132	135	138	141	144	147	150	3
4	164	168	172	176	180	184	188	192	196	200	4
5	205	210	215	220	225	230	235	240	245	250	5
6	246	252	258	264	270	276	282	288	294	300	6
7	287	294	301	308	315	322	329	336	343	350	7
8	328	336	344	352	360	368	376	384	392	400	8
9	369	378	387	396	405	414	423	432	441	450	9
10	410	420	430	440	450	460	470	480	490	500	10
11	451	462	473	484	495	506	517	528	539	550	11
12	492	504	516	528	540	552	564	576	588	600	12
13	533	546	559	572	585	598	611	624	637	650	13
14	574	588	602	616	630	644	658	672	686	700	14
15	615	630	645	660	675	690	705	720	735	750	15
16	656	672	688	704	720	736	752	768	784	800	16
17	697	714	731	748	765	782	799	816	833	850	17
18	738	756	774	792	810	828	846	864	882	900	18
19	779	798	817	836	855	874	893	912	931	950	19
20	820	840	860	880	900	920	940	960	980	1000	20
21	861	882	903	924	945	966	987	1008	1029	1050	21
22	902	924	946	968	990	1012	1034	1056	1078	1100	22
23	943	966	989	1012	1035	1058	1081	1104	1127	1150	23
24	984	1008	1032	1056	1080	1104	1128	1152	1176	1200	24
25	1025	1050	1075	1100	1125	1150	1175	1200	1225	1250	25
26	1066	1092	1118	1144	1170	1196	1222	1248	1274	1300	26
27	1107	1134	1161	1188	1215	1242	1269	1296	1323	1350	27
28	1148	1176	1204	1232	1260	1288	1316	1344	1372	1400	28
29	1189	1218	1247	1276	1305	1334	1363	1392	1421	1450	29
30	1230	1260	1290	1320	1350	1380	1410	1440	1470	1500	30
31	1271	1302	1333	1364	1395	1426	1457	1488	1519	1550	31
32	1312	1344	1376	1408	1440	1472	1504	1536	1568	1600	32
33	1353	1386	1419	1452	1485	1518	1551	1584	1617	1650	33
34	1394	1428	1462	1496	1530	1564	1598	1632	1666	1700	34
35	1435	1470	1505	1540	1575	1610	1645	1680	1715	1750	35
36	1476	1512	1548	1584	1620	1656	1692	1728	1764	1800	36
37	1517	1554	1591	1628	1665	1702	1739	1776	1813	1850	37
38	1558	1596	1634	1672	1710	1748	1786	1824	1862	1900	38
39	1599	1638	1677	1716	1755	1794	1833	1872	1911	1950	39
40	1640	1680	1720	1760	1800	1840	1880	1920	1960	2000	40
41	1681	1722	1763	1804	1845	1886	1927	1968	2009	2050	41
42	1722	1764	1806	1848	1890	1932	1974	2016	2058	2100	42
43	1763	1806	1849	1892	1935	1978	2021	2064	2107	2150	43
44	1804	1848	1892	1936	1980	2024	2068	2112	2156	2200	44
45	1845	1890	1935	1980	2025	2070	2115	2160	2205	2250	45
46	1886	1932	1978	2024	2070	2116	2162	2208	2254	2300	46
47	1927	1974	2021	2068	2115	2162	2209	2256	2303	2350	47
48	1968	2016	2064	2112	2160	2208	2256	2304	2352	2400	48
49	2009	2058	2107	2156	2205	2254	2303	2352	2401	2450	49
50	2050	2100	2150	2200	2250	2300	2350	2400	2450	2500	50
	41	42	43	44	45	46	47	48	49	50	

Multiplication Table

511

	41	42	43	44	45	46	47	48	49	50	
51	2091	2142	2193	2244	2295	2346	2397	2448	2499	2550	51
52	2132	2184	2236	2288	2340	2392	2444	2496	2548	2600	52
53	2173	2226	2279	2332	2385	2438	2491	2544	2597	2650	53
54	2214	2268	2322	2376	2430	2484	2538	2592	2646	2700	54
55	2255	2310	2365	2420	2475	2530	2585	2640	2695	2750	55
56	2296	2352	2408	2464	2520	2576	2632	2688	2744	2800	56
57	2337	2394	2451	2508	2565	2622	2679	2736	2793	2850	57
58	2378	2436	2494	2552	2610	2668	2726	2784	2842	2900	58
59	2419	2478	2537	2596	2655	2714	2773	2832	2891	2950	59
60	2460	2520	2580	2640	2700	2760	2820	2880	2940	3000	60
61	2501	2562	2623	2684	2745	2806	2867	2928	2989	3050	61
62	2542	2604	2666	2728	2790	2852	2914	2976	3038	3100	62
63	2583	2646	2709	2772	2835	2898	2961	3024	3087	3150	63
64	2624	2688	2752	2816	2880	2944	3008	3072	3136	3200	64
65	2665	2730	2795	2860	2925	2990	3055	3120	3185	3250	65
66	2706	2772	2838	2904	2970	3036	3102	3168	3234	3300	66
67	2747	2814	2881	2948	3015	3082	3149	3216	3283	3350	67
68	2788	2856	2924	2992	3060	3128	3196	3264	3332	3400	68
69	2829	2898	2967	3036	3105	3174	3243	3312	3381	3450	69
70	2870	2940	3010	3080	3150	3220	3290	3360	3430	3500	70
71	2911	2982	3053	3124	3195	3266	3337	3408	3479	3550	71
72	2952	3024	3096	3168	3240	3312	3384	3456	3528	3600	72
73	2993	3066	3139	3212	3285	3358	3431	3504	3577	3650	73
74	3034	3108	3182	3256	3330	3404	3478	3552	3626	3700	74
75	3075	3150	3225	3300	3375	3450	3525	3600	3675	3750	75
76	3116	3192	3268	3344	3420	3496	3572	3648	3724	3800	76
77	3157	3234	3311	3388	3465	3542	3619	3696	3773	3850	77
78	3198	3276	3354	3432	3510	3588	3666	3744	3822	3900	78
79	3239	3318	3397	3476	3555	3634	3713	3792	3871	3950	79
80	3280	3360	3440	3520	3600	3680	3760	3840	3920	4000	80
81	3321	3402	3483	3564	3645	3726	3807	3888	3969	4050	81
82	3362	3444	3526	3608	3690	3772	3854	3936	4018	4100	82
83	3403	3486	3569	3652	3735	3818	3901	3984	4067	4150	83
84	3444	3528	3612	3696	3780	3864	3948	4032	4116	4200	84
85	3485	3570	3655	3740	3825	3910	3995	4080	4165	4250	85
86	3526	3612	3698	3784	3870	3956	4042	4128	4214	4300	86
87	3567	3654	3741	3828	3915	4002	4089	4176	4263	4350	87
88	3608	3696	3784	3872	3960	4048	4136	4224	4312	4400	88
89	3649	3738	3827	3916	4005	4094	4183	4272	4361	4450	89
90	3690	3780	3870	3960	4050	4140	4230	4320	4410	4500	90
91	3731	3822	3913	4004	4095	4186	4277	4368	4459	4550	91
92	3772	3864	3956	4048	4140	4232	4324	4416	4508	4600	92
93	3813	3906	3999	4092	4185	4278	4371	4464	4557	4650	93
94	3854	3948	4042	4136	4230	4324	4418	4512	4606	4700	94
95	3895	3990	4085	4180	4275	4370	4465	4560	4655	4750	95
96	3936	4032	4128	4224	4320	4416	4512	4608	4704	4800	96
97	3977	4074	4171	4268	4365	4462	4559	4656	4753	4850	97
98	4018	4116	4214	4312	4410	4508	4606	4704	4802	4900	98
99	4059	4158	4257	4356	4455	4554	4653	4752	4851	4950	99
100	4100	4200	4300	4400	4500	4600	4700	4800	4900	5000	100
	41	42	43	44	45	46	47	48	49	50	

	51	52	53	54	55	56	57	58	59	60	
1	51	52	53	54	55	56	57	58	59	60	1
2	102	104	106	108	110	112	114	116	118	120	2
3	153	156	159	162	165	168	171	174	177	180	3
4	204	208	212	216	220	224	228	232	236	240	4
5	255	260	265	270	275	280	285	290	295	300	5
6	306	312	318	324	330	336	342	348	354	360	6
7	357	364	371	378	385	392	399	406	413	420	7
8	408	416	424	432	440	448	456	464	472	480	8
9	459	468	477	486	495	504	513	522	531	540	9
10	510	520	530	540	550	560	570	580	590	600	10
11	561	572	583	594	605	616	627	638	649	660	11
12	612	624	636	648	660	672	684	696	708	720	12
13	663	676	689	702	715	728	741	754	767	780	13
14	714	728	742	756	770	784	798	812	826	840	14
15	765	780	795	810	825	840	855	870	885	900	15
16	816	832	848	864	880	896	912	928	944	960	16
17	867	884	901	918	935	952	969	986	1003	1020	17
18	918	936	954	972	990	1008	1026	1044	1062	1080	18
19	969	988	1007	1026	1045	1064	1083	1102	1121	1140	19
20	1020	1040	1060	1080	1100	1120	1140	1160	1180	1200	20
21	1071	1092	1113	1134	1155	1176	1197	1218	1239	1260	21
22	1122	1144	1166	1188	1210	1232	1254	1276	1298	1320	22
23	1173	1196	1219	1242	1265	1288	1311	1334	1357	1380	23
24	1224	1248	1272	1296	1320	1344	1368	1392	1416	1440	24
25	1275	1300	1325	1350	1375	1400	1425	1450	1475	1500	25
26	1326	1352	1378	1404	1430	1456	1482	1508	1534	1560	26
27	1377	1404	1431	1458	1485	1512	1539	1566	1593	1620	27
28	1428	1456	1484	1512	1540	1568	1596	1624	1652	1680	28
29	1479	1508	1537	1566	1595	1624	1653	1682	1711	1740	29
30	1530	1560	1590	1620	1650	1680	1710	1740	1770	1800	30
31	1581	1612	1643	1674	1705	1736	1767	1798	1829	1860	31
32	1632	1664	1696	1728	1760	1792	1824	1856	1888	1920	32
33	1683	1716	1749	1782	1815	1848	1881	1914	1947	1980	33
34	1734	1768	1802	1836	1870	1904	1938	1972	2006	2040	34
35	1785	1820	1855	1890	1925	1960	1995	2030	2065	2100	35
36	1836	1872	1908	1944	1980	2016	2052	2088	2124	2160	36
37	1887	1924	1961	1998	2035	2072	2109	2146	2183	2220	37
38	1938	1976	2014	2052	2090	2128	2166	2204	2242	2280	38
39	1989	2028	2067	2106	2145	2184	2223	2262	2301	2340	39
40	2040	2080	2120	2160	2200	2240	2280	2320	2360	2400	40
41	2091	2132	2173	2214	2255	2296	2337	2378	2419	2460	41
42	2142	2184	2226	2268	2310	2352	2394	2436	2478	2520	42
43	2193	2236	2279	2322	2365	2408	2451	2494	2537	2580	43
44	2244	2288	2332	2376	2420	2464	2508	2552	2596	2640	44
45	2295	2340	2385	2430	2475	2520	2565	2610	2655	2700	45
46	2346	2392	2438	2484	2530	2576	2622	2668	2714	2760	46
47	2397	2444	2491	2538	2585	2632	2679	2726	2773	2820	47
48	2448	2496	2544	2592	2640	2688	2736	2784	2832	2880	48
49	2499	2548	2597	2646	2695	2744	2793	2842	2891	2940	49
50	2550	2600	2650	2700	2750	2800	2850	2900	2950	3000	50
	51	52	53	54	55	56	57	58	59	60	

Multiplication Table

513

	51	52	53	54	55	56	57	58	59	60	
51	2601	2652	2703	2754	2805	2856	2907	2958	3009	3060	51
52	2652	2704	2756	2808	2860	2912	2964	3016	3068	3120	52
53	2703	2756	2809	2862	2915	2968	3021	3074	3127	3180	53
54	2754	2808	2862	2916	2970	3024	3078	3132	3186	3240	54
55	2805	2860	2915	2970	3025	3080	3135	3190	3245	3300	55
56	2856	2912	2968	3024	3080	3136	3192	3248	3304	3360	56
57	2907	2964	3021	3078	3135	3192	3249	3306	3363	3420	57
58	2958	3016	3074	3132	3190	3248	3306	3364	3422	3480	58
59	3009	3068	3127	3186	3245	3304	3363	3422	3481	3540	59
60	3060	3120	3180	3240	3300	3360	3420	3480	3540	3600	60
61	3111	3172	3233	3294	3355	3416	3477	3538	3599	3660	61
62	3162	3224	3286	3348	3410	3472	3534	3596	3658	3720	62
63	3213	3276	3339	3402	3465	3528	3591	3654	3717	3780	63
64	3264	3328	3392	3456	3520	3584	3648	3712	3776	3840	64
65	3315	3380	3445	3510	3575	3640	3705	3770	3835	3900	65
66	3366	3432	3498	3564	3630	3696	3762	3828	3894	3960	66
67	3417	3484	3551	3618	3685	3752	3819	3886	3953	4020	67
68	3468	3536	3604	3672	3740	3808	3876	3944	4012	4080	68
69	3519	3588	3657	3726	3795	3864	3933	4002	4071	4140	69
70	3570	3640	3710	3780	3850	3920	3990	4060	4130	4200	70
71	3621	3692	3763	3834	3905	3976	4047	4118	4189	4260	71
72	3672	3744	3816	3888	3960	4032	4104	4176	4248	4320	72
73	3723	3796	3869	3942	4015	4088	4161	4234	4307	4380	73
74	3774	3848	3922	3996	4070	4144	4218	4292	4366	4440	74
75	3825	3900	3975	4050	4125	4200	4275	4350	4425	4500	75
76	3876	3952	4028	4104	4180	4256	4332	4408	4484	4560	76
77	3927	4004	4081	4158	4235	4312	4389	4466	4543	4620	77
78	3978	4056	4134	4212	4290	4368	4446	4524	4602	4680	78
79	4029	4108	4187	4266	4345	4424	4503	4582	4661	4740	79
80	4080	4160	4240	4320	4400	4480	4560	4640	4720	4800	80
81	4131	4212	4293	4374	4455	4536	4617	4698	4779	4860	81
82	4182	4264	4346	4428	4510	4592	4674	4756	4838	4920	82
83	4233	4316	4399	4482	4565	4648	4731	4814	4897	4980	83
84	4284	4368	4452	4536	4620	4704	4788	4872	4956	5040	84
85	4335	4420	4505	4590	4675	4760	4845	4930	5015	5100	85
86	4386	4472	4558	4644	4730	4816	4902	4988	5074	5160	86
87	4437	4524	4611	4698	4785	4872	4959	5046	5133	5220	87
88	4488	4576	4664	4752	4840	4928	5016	5104	5192	5280	88
89	4539	4628	4717	4806	4895	4984	5073	5162	5251	5340	89
90	4590	4680	4770	4860	4950	5040	5130	5220	5310	5400	90
91	4641	4732	4823	4914	5005	5096	5187	5278	5369	5460	91
92	4692	4784	4876	4968	5060	5152	5244	5336	5428	5520	92
93	4743	4836	4929	5022	5115	5208	5301	5394	5487	5580	93
94	4794	4888	4982	5076	5170	5264	5358	5452	5546	5640	94
95	4845	4940	5035	5130	5225	5320	5415	5510	5605	5700	95
96	4896	4992	5088	5184	5280	5376	5472	5568	5664	5760	96
97	4947	5044	5141	5238	5335	5432	5529	5626	5723	5820	97
98	4998	5096	5194	5292	5390	5488	5586	5684	5782	5880	98
99	5049	5148	5247	5346	5445	5544	5643	5742	5841	5940	99
100	5100	5200	5300	5400	5500	5600	5700	5800	5900	6000	100
	51	52	53	54	55	56	57	58	59	60	

	61	62	63	64	65	66	67	68	69	70	
1	61	62	63	64	65	66	67	68	69	70	1
2	122	124	126	128	130	132	134	136	138	140	2
3	183	186	189	192	195	198	201	204	207	210	3
4	244	248	252	256	260	264	268	272	276	280	4
5	305	310	315	320	325	330	335	340	345	350	5
6	366	372	378	384	390	396	402	408	414	420	6
7	427	434	441	448	455	462	469	476	483	490	7
8	488	496	504	512	520	528	536	544	552	560	8
9	549	558	567	576	585	594	603	612	621	630	9
10	610	620	630	640	650	660	670	680	690	700	10
11	671	682	693	704	715	726	737	748	759	770	11
12	732	744	756	768	780	792	804	816	828	840	12
13	793	806	819	832	845	858	871	884	897	910	13
14	854	868	882	896	910	924	938	952	966	980	14
15	915	930	945	960	975	990	1005	1020	1035	1050	15
16	976	992	1008	1024	1040	1056	1072	1088	1104	1120	16
17	1037	1054	1071	1088	1105	1122	1139	1156	1173	1190	17
18	1098	1116	1134	1152	1170	1188	1206	1224	1242	1260	18
19	1159	1178	1197	1216	1235	1254	1273	1292	1311	1330	19
20	1220	1240	1260	1280	1300	1320	1340	1360	1380	1400	20
21	1281	1302	1323	1344	1365	1386	1407	1428	1449	1470	21
22	1342	1364	1386	1408	1430	1452	1474	1496	1518	1540	22
23	1403	1426	1449	1472	1495	1518	1541	1564	1587	1610	23
24	1464	1488	1512	1536	1560	1584	1608	1632	1656	1680	24
25	1525	1550	1575	1600	1625	1650	1675	1700	1725	1750	25
26	1586	1612	1638	1664	1690	1716	1742	1768	1794	1820	26
27	1647	1674	1701	1728	1755	1782	1809	1836	1863	1890	27
28	1708	1736	1764	1792	1820	1848	1876	1904	1932	1960	28
29	1769	1798	1827	1856	1885	1914	1943	1972	2001	2030	29
30	1830	1860	1890	1920	1950	1980	2010	2040	2070	2100	30
31	1891	1922	1953	1984	2015	2046	2077	2108	2139	2170	31
32	1952	1984	2016	2048	2080	2112	2144	2176	2208	2240	32
33	2013	2046	2079	2112	2145	2178	2211	2244	2277	2310	33
34	2074	2108	2142	2176	2210	2244	2278	2312	2346	2380	34
35	2135	2170	2205	2240	2275	2310	2345	2380	2415	2450	35
36	2196	2232	2268	2304	2340	2376	2412	2448	2484	2520	36
37	2257	2294	2331	2368	2405	2442	2479	2516	2553	2590	37
38	2318	2356	2394	2432	2470	2508	2546	2584	2622	2660	38
39	2379	2418	2457	2496	2535	2574	2613	2652	2691	2730	39
40	2440	2480	2520	2560	2600	2640	2680	2720	2760	2800	40
41	2501	2542	2583	2624	2665	2706	2747	2788	2829	2870	41
42	2562	2604	2646	2688	2730	2772	2814	2856	2898	2940	42
43	2623	2666	2709	2752	2795	2838	2881	2924	2967	3010	43
44	2684	2728	2772	2816	2860	2904	2948	2992	3036	3080	44
45	2745	2790	2835	2880	2925	2970	3015	3060	3105	3150	45
46	2806	2852	2898	2944	2990	3036	3082	3128	3174	3220	46
47	2867	2914	2961	3008	3055	3102	3149	3196	3243	3290	47
48	2928	2976	3024	3072	3120	3168	3216	3264	3312	3360	48
49	2989	3038	3087	3136	3185	3234	3283	3332	3381	3430	49
50	3050	3100	3150	3200	3250	3300	3350	3400	3450	3500	50
	61	62	63	64	65	66	67	68	69	70	

Multiplication Table

515

	61	62	63	64	65	66	67	68	69	70	
51	3111	3162	3213	3264	3315	3366	3417	3468	3519	3570	51
52	3172	3224	3276	3328	3380	3432	3484	3536	3588	3640	52
53	3233	3286	3339	3392	3445	3498	3551	3604	3657	3710	53
54	3294	3348	3402	3456	3510	3564	3618	3672	3726	3780	54
55	3355	3410	3465	3520	3575	3630	3685	3740	3795	3850	55
56	3416	3472	3528	3584	3640	3696	3752	3808	3864	3920	56
57	3477	3534	3591	3648	3705	3762	3819	3876	3933	3990	57
58	3538	3596	3654	3712	3770	3828	3886	3944	4002	4060	58
59	3599	3658	3717	3776	3835	3894	3953	4012	4071	4130	59
60	3660	3720	3780	3840	3900	3960	4020	4080	4140	4200	60
61	3721	3782	3843	3904	3965	4026	4087	4148	4209	4270	61
62	3782	3844	3906	3968	4030	4092	4154	4216	4278	4340	62
63	3843	3906	3969	4032	4095	4158	4221	4284	4347	4410	63
64	3904	3968	4032	4096	4160	4224	4288	4352	4416	4480	64
65	3965	4030	4095	4160	4225	4290	4355	4420	4485	4550	65
66	4026	4092	4158	4224	4290	4356	4422	4488	4554	4620	66
67	4087	4154	4221	4288	4355	4422	4489	4556	4623	4690	67
68	4148	4216	4284	4352	4420	4488	4556	4624	4692	4760	68
69	4209	4278	4347	4416	4485	4554	4623	4692	4761	4830	69
70	4270	4340	4410	4480	4550	4620	4690	4760	4830	4900	70
71	4331	4402	4473	4544	4615	4686	4757	4828	4899	4970	71
72	4392	4464	4536	4608	4680	4752	4824	4896	4968	5040	72
73	4453	4526	4599	4672	4745	4818	4891	4964	5037	5110	73
74	4514	4588	4662	4736	4810	4884	4958	5032	5106	5180	74
75	4575	4650	4725	4800	4875	4950	5025	5100	5175	5250	75
76	4636	4712	4788	4864	4940	5016	5092	5168	5244	5320	76
77	4697	4774	4851	4928	5005	5082	5159	5236	5313	5390	77
78	4758	4836	4914	4992	5070	5148	5226	5304	5382	5460	78
79	4819	4898	4977	5056	5135	5214	5293	5372	5451	5530	79
80	4880	4960	5040	5120	5200	5280	5360	5440	5520	5600	80
81	4941	5022	5103	5184	5265	5346	5427	5508	5589	5670	81
82	5002	5084	5166	5248	5330	5412	5494	5576	5658	5740	82
83	5063	5146	5229	5312	5395	5478	5561	5644	5727	5810	83
84	5124	5208	5292	5376	5460	5544	5628	5712	5796	5880	84
85	5185	5270	5355	5440	5525	5610	5695	5780	5865	5950	85
86	5246	5332	5418	5504	5590	5676	5762	5848	5934	6020	86
87	5307	5394	5481	5568	5655	5742	5829	5916	6003	6090	87
88	5368	5456	5544	5632	5720	5808	5896	5984	6072	6160	88
89	5429	5518	5607	5696	5785	5874	5963	6052	6141	6230	89
90	5490	5580	5670	5760	5850	5940	6030	6120	6210	6300	90
91	5551	5642	5733	5824	5915	6006	6097	6188	6279	6370	91
92	5612	5704	5796	5888	5980	6072	6164	6256	6348	6440	92
93	5673	5766	5859	5952	6045	6138	6231	6324	6417	6510	93
94	5734	5828	5922	6016	6110	6204	6298	6392	6486	6580	94
95	5795	5890	5985	6080	6175	6270	6365	6460	6555	6650	95
96	5856	5952	6048	6144	6240	6336	6432	6528	6624	6720	96
97	5917	6014	6111	6208	6305	6402	6499	6596	6693	6790	97
98	5978	6076	6174	6272	6370	6468	6566	6664	6762	6860	98
99	6039	6138	6237	6336	6435	6534	6633	6732	6831	6930	99
100	6100	6200	6300	6400	6500	6600	6700	6800	6900	7000	100
	61	62	63	64	65	66	67	68	69	70	

	71	72	73	74	75	76	77	78	79	80	
1	71	72	73	74	75	76	77	78	79	80	1
2	142	144	146	148	150	152	154	156	158	160	2
3	213	216	219	222	225	228	231	234	237	240	3
4	284	288	292	296	300	304	308	312	316	320	4
5	355	360	365	370	375	380	385	390	395	400	5
6	426	432	438	444	450	456	462	468	474	480	6
7	497	504	511	518	525	532	539	546	553	560	7
8	568	576	584	592	600	608	616	624	632	640	8
9	639	648	657	666	675	684	693	702	711	720	9
10	710	720	730	740	750	760	770	780	790	800	10
11	781	792	803	814	825	836	847	858	869	880	11
12	852	864	876	888	900	912	924	936	948	960	12
13	923	936	949	962	975	988	1001	1014	1027	1040	13
14	994	1008	1022	1036	1050	1064	1078	1092	1106	1120	14
15	1065	1080	1095	1110	1125	1140	1155	1170	1185	1200	15
16	1136	1152	1168	1184	1200	1216	1232	1248	1264	1280	16
17	1207	1224	1241	1258	1275	1292	1309	1326	1343	1360	17
18	1278	1296	1314	1332	1350	1368	1386	1404	1422	1440	18
19	1349	1368	1387	1406	1425	1444	1463	1482	1501	1520	19
20	1420	1440	1460	1480	1500	1520	1540	1560	1580	1600	20
21	1491	1512	1533	1554	1575	1596	1617	1638	1659	1680	21
22	1562	1584	1606	1628	1650	1672	1694	1716	1738	1760	22
23	1633	1656	1679	1702	1725	1748	1771	1794	1817	1840	23
24	1704	1728	1752	1776	1800	1824	1848	1872	1896	1920	24
25	1775	1800	1825	1850	1875	1900	1925	1950	1975	2000	25
26	1846	1872	1898	1924	1950	1976	2002	2028	2054	2080	26
27	1917	1944	1971	1998	2025	2052	2079	2106	2133	2160	27
28	1988	2016	2044	2072	2100	2128	2156	2184	2212	2240	28
29	2059	2088	2117	2146	2175	2204	2233	2262	2291	2320	29
30	2130	2160	2190	2220	2250	2280	2310	2340	2370	2400	30
31	2201	2232	2263	2294	2325	2356	2387	2418	2449	2480	31
32	2272	2304	2336	2368	2400	2432	2464	2496	2528	2560	32
33	2343	2376	2409	2442	2475	2508	2541	2574	2607	2640	33
34	2414	2448	2482	2516	2550	2584	2618	2652	2686	2720	34
35	2485	2520	2555	2590	2625	2660	2695	2730	2765	2800	35
36	2556	2592	2628	2664	2700	2736	2772	2808	2844	2880	36
37	2627	2664	2701	2738	2775	2812	2849	2886	2923	2960	37
38	2698	2736	2774	2812	2850	2888	2926	2964	3002	3040	38
39	2769	2808	2847	2886	2925	2964	3003	3042	3081	3120	39
40	2840	2880	2920	2960	3000	3040	3080	3120	3160	3200	40
41	2911	2952	2993	3034	3075	3116	3157	3198	3239	3280	41
42	2982	3024	3066	3108	3150	3192	3234	3276	3318	3360	42
43	3053	3096	3139	3182	3225	3268	3311	3354	3397	3440	43
44	3124	3168	3212	3256	3300	3344	3388	3432	3476	3520	44
45	3195	3240	3285	3330	3375	3420	3465	3510	3555	3600	45
46	3266	3312	3358	3404	3450	3496	3542	3588	3634	3680	46
47	3337	3384	3431	3478	3525	3572	3619	3666	3713	3760	47
48	3408	3456	3504	3552	3600	3648	3696	3744	3792	3840	48
49	3479	3528	3577	3626	3675	3724	3773	3822	3871	3920	49
50	3550	3600	3650	3700	3750	3800	3850	3900	3950	4000	50
	71	72	73	74	75	76	77	78	79	80	

Multiplication Table

517

	71	72	73	74	75	76	77	78	79	80	
51	3621	3672	3723	3774	3825	3876	3927	3978	4029	4080	51
52	3692	3744	3796	3848	3900	3952	4004	4056	4108	4160	52
53	3763	3816	3869	3922	3975	4028	4081	4134	4187	4240	53
54	3834	3888	3942	3996	4050	4104	4158	4212	4266	4320	54
55	3905	3960	4015	4070	4125	4180	4235	4290	4345	4400	55
56	3976	4032	4088	4144	4200	4256	4312	4368	4424	4480	56
57	4047	4104	4161	4218	4275	4332	4389	4446	4503	4560	57
58	4118	4176	4234	4292	4350	4408	4466	4524	4582	4640	58
59	4189	4248	4307	4366	4425	4484	4543	4602	4661	4720	59
60	4260	4320	4380	4440	4500	4560	4620	4680	4740	4800	60
61	4331	4392	4453	4514	4575	4636	4697	4758	4819	4880	61
62	4402	4464	4526	4588	4650	4712	4774	4836	4898	4960	62
63	4473	4536	4599	4662	4725	4788	4851	4914	4977	5040	63
64	4544	4608	4672	4736	4800	4864	4928	4992	5056	5120	64
65	4615	4680	4745	4810	4875	4940	5005	5070	5135	5200	65
66	4686	4752	4818	4884	4950	5016	5082	5148	5214	5280	66
67	4757	4824	4891	4958	5025	5092	5159	5226	5293	5360	67
68	4828	4896	4964	5032	5100	5168	5236	5304	5372	5440	68
69	4899	4968	5037	5106	5175	5244	5313	5382	5451	5520	69
70	4970	5040	5110	5180	5250	5320	5390	5460	5530	5600	70
71	5041	5112	5183	5254	5325	5396	5467	5538	5609	5680	71
72	5112	5184	5256	5328	5400	5472	5544	5616	5688	5760	72
73	5183	5256	5329	5402	5475	5548	5621	5694	5767	5840	73
74	5254	5328	5402	5476	5550	5624	5698	5772	5846	5920	74
75	5325	5400	5475	5550	5625	5700	5775	5850	5925	6000	75
76	5396	5472	5548	5624	5700	5776	5852	5928	6004	6080	76
77	5467	5544	5621	5698	5775	5852	5929	6006	6083	6160	77
78	5538	5616	5694	5772	5850	5928	6006	6084	6162	6240	78
79	5609	5688	5767	5846	5925	6004	6083	6162	6241	6320	79
80	5680	5760	5840	5920	6000	6080	6160	6240	6320	6400	80
81	5751	5832	5913	5994	6075	6156	6237	6318	6399	6480	81
82	5822	5904	5986	6068	6150	6232	6314	6396	6478	6560	82
83	5893	5976	6059	6142	6225	6308	6391	6474	6557	6640	83
84	5964	6048	6132	6216	6300	6384	6468	6552	6636	6720	84
85	6035	6120	6205	6290	6375	6460	6545	6630	6715	6800	85
86	6106	6192	6278	6364	6450	6536	6622	6708	6794	6880	86
87	6177	6264	6351	6438	6525	6612	6699	6786	6873	6960	87
88	6248	6336	6424	6512	6600	6688	6776	6864	6952	7040	88
89	6319	6408	6497	6586	6675	6764	6853	6942	7031	7120	89
90	6390	6480	6570	6660	6750	6840	6930	7020	7110	7200	90
91	6461	6552	6643	6734	6825	6916	7007	7098	7189	7280	91
92	6532	6624	6716	6808	6900	6992	7084	7176	7268	7360	92
93	6603	6696	6789	6882	6975	7068	7161	7254	7347	7440	93
94	6674	6768	6862	6956	7050	7144	7238	7332	7426	7520	94
95	6745	6840	6935	7030	7125	7220	7315	7410	7505	7600	95
96	6816	6912	7008	7104	7200	7296	7392	7488	7584	7680	96
97	6887	6984	7081	7178	7275	7372	7469	7566	7663	7760	97
98	6958	7056	7154	7252	7350	7448	7546	7644	7742	7840	98
99	7029	7128	7227	7326	7425	7524	7623	7722	7821	7920	99
100	7100	7200	7300	7400	7500	7600	7700	7800	7900	8000	100
	71	72	73	74	75	76	77	78	79	80	

	81	82	83	84	85	86	87	88	89	90	
1	81	82	83	84	85	86	87	88	89	90	1
2	162	164	166	168	170	172	174	176	178	180	2
3	243	246	249	252	255	258	261	264	267	270	3
4	324	328	332	336	340	344	348	352	356	360	4
5	405	410	415	420	425	430	435	440	445	450	5
6	486	492	498	504	510	516	522	528	534	540	6
7	567	574	581	588	595	602	609	616	623	630	7
8	648	656	664	672	680	688	696	704	712	720	8
9	729	738	747	756	765	774	783	792	801	810	9
10	810	820	830	840	850	860	870	880	890	900	10
11	891	902	913	924	935	946	957	968	979	990	11
12	972	984	996	1008	1020	1032	1044	1056	1068	1080	12
13	1053	1066	1079	1092	1105	1118	1131	1144	1157	1170	13
14	1184	1148	1162	1176	1190	1204	1218	1232	1246	1260	14
15	1215	1230	1245	1260	1275	1290	1305	1320	1335	1350	15
16	1296	1312	1328	1344	1360	1376	1392	1408	1424	1440	16
17	1377	1394	1411	1428	1445	1462	1479	1496	1513	1530	17
18	1458	1476	1494	1512	1530	1548	1566	1584	1602	1620	18
19	1539	1558	1577	1596	1615	1634	1653	1672	1691	1710	19
20	1620	1640	1660	1680	1700	1720	1740	1760	1780	1800	20
21	1701	1722	1743	1764	1785	1806	1827	1848	1869	1890	21
22	1782	1804	1826	1848	1870	1892	1914	1936	1958	1980	22
23	1863	1886	1909	1932	1955	1978	2001	2024	2047	2070	23
24	1944	1968	1992	2016	2040	2064	2088	2112	2136	2160	24
25	2025	2050	2075	2100	2125	2150	2175	2200	2225	2250	25
26	2106	2132	2158	2184	2210	2236	2262	2288	2314	2340	26
27	2187	2214	2241	2268	2295	2322	2349	2376	2403	2430	27
28	2268	2296	2324	2352	2380	2408	2436	2464	2492	2520	28
29	2349	2378	2407	2436	2465	2494	2523	2552	2581	2610	29
30	2430	2460	2490	2520	2550	2580	2610	2640	2670	2700	30
31	2511	2542	2573	2604	2635	2666	2697	2728	2759	2790	31
32	2592	2624	2656	2688	2720	2752	2784	2816	2848	2880	32
33	2673	2706	2739	2772	2805	2838	2871	2904	2937	2970	33
34	2754	2788	2822	2856	2890	2924	2958	2992	3026	3060	34
35	2835	2870	2905	2940	2975	3010	3045	3080	3115	3150	35
36	2916	2952	2988	3024	3060	3096	3132	3168	3204	3240	36
37	2997	3034	3071	3108	3145	3182	3219	3256	3293	3330	37
38	3078	3116	3154	3192	3230	3268	3306	3344	3382	3420	38
39	3159	3198	3237	3276	3315	3354	3393	3432	3471	3510	39
40	3240	3280	3320	3360	3400	3440	3480	3520	3560	3600	40
41	3321	3362	3403	3444	3485	3526	3567	3608	3649	3690	41
42	3402	3444	3486	3528	3570	3612	3654	3696	3738	3780	42
43	3483	3526	3569	3612	3655	3698	3741	3784	3827	3870	43
44	3564	3608	3652	3696	3740	3784	3828	3872	3916	3960	44
45	3645	3690	3735	3780	3825	3870	3915	3960	4005	4050	45
46	3726	3772	3818	3864	3910	3956	4002	4048	4094	4140	46
47	3807	3854	3901	3948	3995	4042	4089	4136	4183	4230	47
48	3888	3936	3984	4032	4080	4128	4176	4224	4272	4320	48
49	3969	4018	4067	4116	4165	4214	4263	4312	4361	4410	49
50	4050	4100	4150	4200	4250	4300	4350	4400	4450	4500	50
	81	82	83	84	85	86	87	88	89	90	

Multiplication Table

519

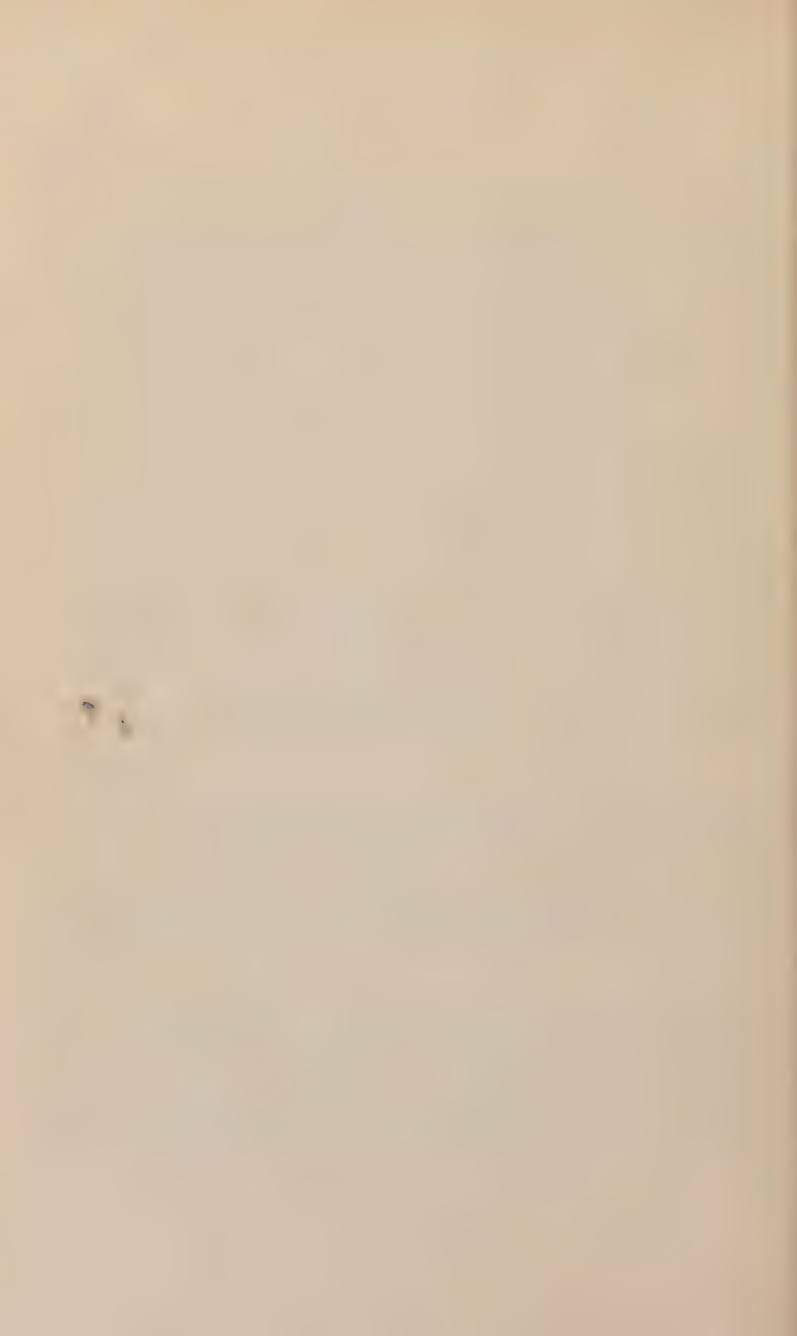
	81	82	83	84	85	86	87	88	89	90	
51	4131	4182	4233	4284	4335	4386	4437	4488	4539	4590	51
52	4212	4264	4316	4368	4420	4472	4524	4576	4628	4680	52
53	4293	4346	4399	4452	4505	4558	4611	4664	4717	4770	53
54	4374	4428	4482	4536	4590	4644	4698	4752	4806	4860	54
55	4455	4510	4565	4620	4675	4730	4785	4840	4895	4950	55
56	4536	4592	4648	4704	4760	4816	4872	4928	4984	5040	56
57	4617	4674	4731	4788	4845	4902	4959	5016	5073	5130	57
58	4698	4756	4814	4872	4930	4988	5046	5104	5162	5220	58
59	4779	4838	4897	4956	5015	5074	5133	5192	5251	5310	59
60	4860	4920	4980	5040	5100	5160	5220	5280	5340	5400	60
61	4941	5002	5063	5124	5185	5246	5307	5368	5429	5490	61
62	5022	5084	5146	5208	5270	5332	5394	5456	5518	5580	62
63	5103	5166	5229	5292	5355	5418	5481	5544	5607	5670	63
64	5184	5248	5312	5376	5440	5504	5568	5632	5696	5760	64
65	5265	5330	5395	5460	5525	5590	5655	5720	5785	5850	65
66	5346	5412	5478	5544	5610	5676	5742	5808	5874	5940	66
67	5427	5494	5561	5628	5695	5762	5829	5896	5963	6030	67
68	5508	5576	5644	5712	5780	5848	5916	5984	6052	6120	68
69	5589	5658	5727	5796	5865	5934	6003	6072	6141	6210	69
70	5670	5740	5810	5880	5950	6020	6090	6160	6230	6300	70
71	5751	5822	5893	5964	6035	6106	6177	6248	6319	6390	71
72	5832	5904	5976	6048	6120	6192	6264	6336	6408	6480	72
73	5913	5986	6059	6132	6205	6278	6351	6424	6497	6570	73
74	5994	6068	6142	6216	6290	6364	6438	6512	6586	6660	74
75	6075	6150	6225	6300	6375	6450	6525	6600	6675	6750	75
76	6156	6232	6308	6384	6460	6536	6612	6688	6764	6840	76
77	6237	6314	6391	6468	6545	6622	6699	6776	6853	6930	77
78	6318	6396	6474	6552	6630	6708	6786	6864	6942	7020	78
79	6399	6478	6557	6636	6715	6794	6873	6952	7031	7110	79
80	6480	6560	6640	6720	6800	6880	6960	7040	7120	7200	80
81	6561	6642	6723	6804	6885	6966	7047	7128	7209	7290	81
82	6642	6724	6806	6888	6970	7052	7134	7216	7298	7380	82
83	6723	6806	6889	6972	7055	7138	7221	7304	7387	7470	83
84	6804	6888	6972	7056	7140	7224	7308	7392	7476	7560	84
85	6885	6970	7055	7140	7225	7310	7395	7480	7565	7650	85
86	6966	7052	7138	7224	7310	7396	7482	7568	7654	7740	86
87	7047	7134	7221	7308	7395	7482	7569	7656	7743	7830	87
88	7128	7216	7304	7392	7480	7568	7656	7744	7832	7920	88
89	7209	7298	7387	7476	7565	7654	7743	7832	7921	8010	89
90	7290	7380	7470	7560	7650	7740	7830	7920	8010	8100	90
91	7371	7462	7553	7644	7735	7826	7917	8008	8099	8190	91
92	7452	7544	7636	7728	7820	7912	8004	8096	8188	8280	92
93	7533	7626	7719	7812	7905	7998	8091	8184	8277	8370	93
94	7614	7708	7802	7896	7990	8084	8178	8272	8366	8460	94
95	7695	7790	7885	7980	8075	8170	8265	8360	8455	8550	95
96	7776	7872	7968	8064	8160	8256	8352	8448	8544	8640	96
97	7857	7954	8051	8148	8245	8342	8439	8536	8633	8730	97
98	7938	8036	8134	8232	8330	8428	8526	8624	8722	8820	98
99	8019	8118	8217	8316	8415	8514	8613	8712	8811	8910	99
100	8100	8200	8300	8400	8500	8600	8700	8800	8900	9000	100
	81	82	83	84	85	86	87	88	89	90	

	91	92	93	94	95	96	97	98	99	100	
1	91	92	93	94	95	96	97	98	99	100	1
2	182	184	186	188	190	192	194	196	198	200	2
3	273	276	279	282	285	288	291	294	297	300	3
4	364	368	372	376	380	384	388	392	396	400	4
5	455	460	465	470	475	480	485	490	495	500	5
6	546	552	558	564	570	576	582	588	594	600	6
7	637	644	651	658	665	672	679	686	693	700	7
8	728	736	744	752	760	768	776	784	792	800	8
9	819	828	837	846	855	864	873	882	891	900	9
10	910	920	930	940	950	960	970	980	990	1000	10
11	1001	1012	1023	1034	1045	1056	1067	1078	1089	1100	11
12	1092	1104	1116	1128	1140	1152	1164	1176	1188	1200	12
13	1183	1196	1209	1222	1235	1248	1261	1274	1287	1300	13
14	1274	1288	1302	1316	1330	1344	1358	1372	1386	1400	14
15	1365	1380	1395	1410	1425	1440	1455	1470	1485	1500	15
16	1456	1472	1488	1504	1520	1536	1552	1568	1584	1600	16
17	1547	1564	1581	1598	1615	1632	1649	1666	1683	1700	17
18	1638	1656	1674	1692	1710	1728	1746	1764	1782	1800	18
19	1729	1748	1767	1786	1805	1824	1843	1862	1881	1900	19
20	1820	1840	1860	1880	1900	1920	1940	1960	1980	2000	20
21	1911	1932	1953	1974	1995	2016	2037	2058	2079	2100	21
22	2002	2024	2046	2068	2090	2112	2134	2156	2178	2200	22
23	2093	2116	2139	2162	2185	2208	2231	2254	2277	2300	23
24	2184	2208	2232	2256	2280	2304	2328	2352	2376	2400	24
25	2275	2300	2325	2350	2375	2400	2425	2450	2475	2500	25
26	2366	2392	2418	2444	2470	2496	2522	2548	2574	2600	26
27	2457	2484	2511	2538	2565	2592	2619	2646	2673	2700	27
28	2548	2576	2604	2632	2660	2688	2716	2744	2772	2800	28
29	2639	2668	2697	2726	2755	2784	2813	2842	2871	2900	29
30	2730	2760	2790	2820	2850	2880	2910	2940	2970	3000	30
31	2821	2852	2883	2914	2945	2976	3007	3038	3069	3100	31
32	2912	2944	2976	3008	3040	3072	3104	3136	3168	3200	32
33	3003	3036	3069	3102	3135	3168	3201	3234	3267	3300	33
34	3094	3128	3162	3196	3230	3264	3298	3332	3366	3400	34
35	3185	3220	3255	3290	3325	3360	3395	3430	3465	3500	35
36	3276	3312	3348	3384	3420	3456	3492	3528	3564	3600	36
37	3367	3404	3441	3478	3515	3552	3589	3626	3663	3700	37
38	3458	3496	3534	3572	3610	3648	3686	3724	3762	3800	38
39	3549	3588	3627	3666	3705	3744	3783	3822	3861	3900	39
40	3640	3680	3720	3760	3800	3840	3880	3920	3960	4000	40
41	3731	3772	3813	3854	3895	3936	3977	4018	4059	4100	41
42	3822	3864	3906	3948	3990	4032	4074	4116	4158	4200	42
43	3913	3956	3999	4042	4085	4128	4171	4214	4257	4300	43
44	4004	4048	4092	4136	4180	4224	4268	4312	4356	4400	44
45	4095	4140	4185	4230	4275	4320	4365	4410	4455	4500	45
46	4186	4232	4278	4324	4370	4416	4462	4508	4554	4600	46
47	4277	4324	4371	4418	4465	4512	4559	4606	4653	4700	47
48	4368	4416	4464	4512	4560	4608	4656	4704	4752	4800	48
49	4459	4508	4557	4606	4655	4704	4753	4802	4851	4900	49
50	4550	4600	4650	4700	4750	4800	4850	4900	4950	5000	50
	91	92	93	94	95	96	97	98	99	100	

Multiplication Table

521

	91	92	93	94	95	96	97	98	99	100	
51	4641	4692	4743	4794	4845	4896	4947	4998	5049	5100	51
52	4732	4784	4836	4888	4940	4992	5044	5096	5148	5200	52
53	4823	4876	4929	4982	5035	5088	5141	5194	5247	5300	53
54	4914	4968	5022	5076	5130	5184	5238	5292	5346	5400	54
55	5005	5060	5115	5170	5225	5280	5335	5390	5445	5500	55
56	5096	5152	5208	5264	5320	5376	5432	5488	5544	5600	56
57	5187	5244	5301	5358	5415	5472	5529	5586	5643	5700	57
58	5278	5336	5394	5452	5510	5568	5626	5684	5742	5800	58
59	5369	5428	5487	5546	5605	5664	5723	5782	5841	5900	59
60	5460	5520	5580	5640	5700	5760	5820	5880	5940	6000	60
61	5551	5612	5673	5734	5795	5856	5917	5978	6039	6100	61
62	5642	5704	5766	5828	5890	5952	6014	6076	6138	6200	62
63	5733	5796	5859	5922	5985	6048	6111	6174	6237	6300	63
64	5824	5888	5952	6016	6080	6144	6208	6272	6336	6400	64
65	5915	5980	6045	6110	6175	6240	6305	6370	6435	6500	65
66	6006	6072	6138	6204	6270	6336	6402	6468	6534	6600	66
67	6097	6164	6231	6298	6365	6432	6499	6566	6633	6700	67
68	6188	6256	6324	6392	6460	6528	6596	6664	6732	6800	68
69	6279	6348	6417	6486	6555	6624	6693	6762	6831	6900	69
70	6370	6440	6510	6580	6650	6720	6790	6860	6930	7000	70
71	6461	6532	6603	6674	6745	6816	6887	6958	7029	7100	71
72	6552	6624	6696	6768	6840	6912	6984	7056	7128	7200	72
73	6643	6716	6789	6862	6935	7008	7081	7154	7227	7300	73
74	6734	6808	6882	6956	7030	7104	7178	7252	7326	7400	74
75	6825	6900	6975	7050	7125	7200	7275	7350	7425	7500	75
76	6916	6992	7068	7144	7220	7296	7372	7448	7524	7600	76
77	7007	7084	7161	7238	7315	7392	7469	7546	7623	7700	77
78	7098	7176	7254	7332	7410	7488	7566	7644	7722	7800	78
79	7189	7268	7347	7426	7505	7584	7663	7742	7821	7900	79
80	7280	7360	7440	7520	7600	7680	7760	7840	7920	8000	80
81	7371	7452	7533	7614	7695	7776	7857	7938	8019	8100	81
82	7462	7544	7626	7708	7790	7872	7954	8036	8118	8200	82
83	7553	7636	7719	7802	7885	7968	8051	8134	8217	8300	83
84	7644	7728	7812	7896	7980	8064	8148	8232	8316	8400	84
85	7735	7820	7905	7990	8075	8160	8245	8330	8415	8500	85
86	7826	7912	7998	8084	8170	8256	8342	8428	8514	8600	86
87	7917	8004	8091	8178	8265	8352	8439	8526	8613	8700	87
88	8008	8096	8184	8272	8360	8448	8536	8624	8712	8800	88
89	8099	8188	8277	8366	8455	8544	8633	8722	8811	8900	89
90	8190	8280	8370	8460	8550	8640	8730	8820	8910	9000	90
91	8281	8372	8463	8554	8645	8736	8827	8918	9009	9100	91
92	8372	8464	8556	8648	8740	8832	8924	9016	9108	9200	92
93	8463	8556	8649	8742	8835	8928	9021	9114	9207	9300	93
94	8554	8648	8742	8836	8930	9024	9118	9212	9306	9400	94
95	8645	8740	8835	8930	9025	9120	9215	9310	9405	9500	95
96	8736	8832	8928	9024	9120	9216	9312	9408	9504	9600	96
97	8827	8924	9021	9118	9215	9312	9409	9506	9603	9700	97
98	8918	9016	9114	9212	9310	9408	9506	9604	9702	9800	98
99	9009	9108	9207	9306	9405	9504	9603	9702	9801	9900	99
100	9100	9200	9300	9400	9500	9600	9700	9800	9900	10000	100
	91	92	93	94	95	96	97	98	99	100	



REFERENCES

No attempt is made here to present a complete bibliography. These titles are intended mainly to indicate the sources of material referred to in the present volume.

1. ANDERSON, L. D. "Estimating Intelligence by Means of Printed Photographs." *Journal of Applied Psychology*, Vol. 5 (1921), pages 152-155.
2. BARTON, ELLA. *Correlations among Motor Abilities*. A.B. Thesis; 1926. (Filed in University of Wisconsin Library.)
3. BILLS, M. A. "Methods for the Selection of Comptometer Operators and Stenographers." *Journal of Applied Psychology*, Vol. 5 (1921), pages 275-288.
4. BINET, ALFRED. *Les révélations de l'écriture d'après un contrôle scientifique*. Paris; 1906. 257 pages.
5. — and SIMON, TH. "Le développement de l'intelligence chez les enfants." *L'Année Psychologique*, Vol. 14 (1908), pages 1-94.
6. BOOK, W. F. "Voluntary Motor Ability of the World's Champion Typists." *Journal of Applied Psychology*, Vol. 8 (1924), pages 283-308.
7. BRIDGES, J. W., and DOLLINGER, M. "The Correlation between Interests and Abilities in College Courses." *Psychological Review*, Vol. 27 (1920), pages 308-314.
8. BROWN, LOIS E. *An Experimental Investigation of the Alleged Relations between Certain Character Traits and Handwriting*. A.B. Thesis; 1921. (Filed in University of Wisconsin Library.)
9. BROWN, W., and THOMSON, G. H. *The Essentials of Mental Measurement*. University Press, Cambridge, England; 1921. 216 pages.
10. BURKS, BARBARA S. "On the Inadequacy of the Partial and Multiple Correlation Technique." *Journal of Educational Psychology*, Vol. 17 (1926), pages 532-540, 625-630.
11. BURTT, HAROLD E. *Principles of Employment Psychology*. Houghton Mifflin Company, Boston; 1926. 568 pages.
12. CADES, ELLIOT. "The Textile Industry in Philadelphia." *Psychological Clinic*, Vol. 15, No. 6 and No. 7 (1924), pages 201-228.
13. CATTELL, J. McKEEN. "Mental Tests and Measurements." *Mind*, Vol. 15 (1890), pages 373-380.
14. CHAPMAN, J. C. *Trade Tests*. Henry Holt & Co., New York; 1921.
15. CLEETON, G. U., and KNIGHT, F. B. "Validity of Character Judgments Based on External Criteria." *Journal of Applied Psychology*, Vol. 8 (1924), pages 215-229.
16. CORRELL, E. F., HOGAN, E. M., and SPENSLEY, A. L. *Aptitude Tests for Department-Store Cashiers*. A.B. Thesis; 1922. (Filed in University of Wisconsin Library.)

17. CRAMPTON, C. W. "The Blood Ptosis Test and Its Use in Experimental Hygiene." *Proceedings of the Society of Experimental Biology and Medicine*, Vol. 12, page 119.
18. DOOLITTLE, M. H. "Adjustment of the Primary Triangulation between Kent Island and Atlanta Base Lines." (Report of the Superintendent.) *Coast and Geodetic Survey* (1878), pages 115-120.
19. EVANS, ALICE L. *The Alleged Relations between the Face and the Character*. A.B. Thesis; 1921. (Filed in University of Wisconsin Library.)
20. FRANKLIN, E. E. "The Permanence of Vocational Interests of Junior High School Pupils." *Johns Hopkins Studies in Education*, No. 8. Johns Hopkins Press, Baltimore; 1924. 68 pages.
21. FREYD, MAX. "The Personalities of the Socially and the Mechanically Inclined." *Psychological Monographs*, Vol. 33, No. 151 (1923). 191 pages.
22. GARNETT, J. C. M. "The Single General Factor in Dissimilar Mental Measurements." *British Journal of Psychology*, Vol. 10 (1920), pages 242-258.
23. GAW, FRANCES. "A Study of Performance Tests." *British Journal of Psychology*, Vol. 15 (1925), pages 374-392.
24. GILBRETH, F. B. and L. M. *Applied Motion Study*. The Macmillan Company, New York; 1919. 220 pages.
25. ——— *Fatigue Study*. The Macmillan Company, New York; 1919. 175 pages.
26. GRIFFITTS, C. H. *Fundamentals of Vocational Psychology*. The Macmillan Company, New York; 1924. 372 pages.
27. HANDSCHIN, C. H. "A Test for Discovering Types of Learning in Language Study." *Modern Language Journal*, Vol. 3 (1918), pages 1-4.
28. HARRIS, J. A. "A Short Method of Calculating the Coefficient of Correlation in the Case of Integral Variables." *Biometrika*, Vol. 7 (1909), page 214.
29. HEIDBREDER, EDNA. "Intelligence and the Height-Weight Ratio." *Journal of Applied Psychology*, Vol. 10, No. 1 (1926), pages 52-62.
30. HENMON, V. A. C. "Air-Service Tests of Aptitude for Flying." *Journal of Applied Psychology*, Vol. 3 (1919), pages 103-109.
31. HOKE, E. R. "The Measurement of Achievement in Shorthand." *Johns Hopkins Studies in Education*, No. 6. Johns Hopkins Press, Baltimore; 1922. 118 pages.
32. ——— *Prognostic Tests of Stenographic Ability*. Gregg Publishing Company, New York.
33. HOLLINGWORTH, H. L. *Vocational Psychology*. D. Appleton & Co., New York; 1916. 308 pages.
34. HULL, CLARK L. "An Automatic Correlation Calculating Machine." *Journal of American Statistical Association*, Vol. 20 (1925), pages 522-531.
35. ——— "An Automatic Machine for Making Multiple Aptitude Forecasts." *Journal of Educational Psychology*, Vol. 16 (1925), pages 593-598.

- 35 a. HULL, CLARK L. "The Joint Yield from Teams of Tests." *Journal of Educational Psychology*, Vol. 14 (1923), pages 396-406.
36. ——— "Prediction Formulæ for Teams of Aptitude Tests." *Journal of Applied Psychology*, Vol. 7 (1923), pages 277-284.
37. ——— "The Computation of Pearson's r from Ranked Data." *Journal of Applied Psychology*, Vol. 6 (1922), pages 385-390.
38. ——— "The Conversion of Test Scores into Series Which Shall Have Assigned Mean and Degree of Dispersion." *Journal of Applied Psychology*, Vol. 6 (1922), pages 298-300.
39. ——— "The Correlation Coefficient and Its Prognostic Significance." *Journal of Educational Research*, Vol. 15 (1927), pages 327-338.
40. ——— "Variability in Amount of Different Traits Possessed by the Individual." *Journal of Educational Psychology*, Vol. 18 (1927), pages 97-104.
41. ——— and MONTGOMERY, R. B. "An Experimental Investigation of Certain Alleged Relations between Character and Handwriting." *Psychological Review*, Vol. 26 (1919), pages 63-75.
42. ——— and LIMP, C. E. "The Differentiation of the Aptitudes of an Individual by Means of Test Batteries." *Journal of Educational Psychology*, Vol. 16 (1925), pages 73-88.
43. KELLEY, TRUMAN L. "A New Method for Determining the Significance of Differences in Intelligence and Achievement Scores." *Journal of Educational Psychology*, Vol. 14 (1923), pages 321-333.
44. ——— *Educational Guidance*. Bureau of Publications, Teachers College, Columbia University, New York; 1914.
45. ——— *Statistical Method*. The Macmillan Company, New York; 1923. 385 pages.
46. KEMBLE, W. F. *Choosing Employees by Test*. Engineering Magazine Company, 381 Fourth Avenue, New York; 1917. 333 pages.
47. KORNHAUSER, A. W., and KINGSBURY, F. A. *Psychological Tests in Business*. University of Chicago Press, Chicago; 1924. 194 pages.
48. KRETSCHMER, E. *Physique and Character* (translated by W. T. H. Sprott). Harcourt, Brace & Co., New York; 1925.
49. LAIRD, DONALD A. *The Psychology of Selecting Men*. McGraw-Hill Book Company, New York; 1925. 274 pages.
50. LIMP, CHARLES E. "The Use of the Regression Equation in Determining the Aptitudes of an Individual." *Journal of Educational Psychology*, Vol. 16 (1925), pages 414-418.
51. LINK, HENRY C. *Employment Psychology*. The Macmillan Company, New York; 1919. 440 pages.
52. McCABE, FLORENCE E. *The Relation between Character Traits and Judgments of Character Based on Photographs*. A.B. Thesis; 1926. (Filed in University of Wisconsin Library.)
53. MCCALL, W. A. *How to Measure in Education*. The Macmillan Company, New York; 1922.

54. MCFARLANE, MARGARET. "A Study of Practical Ability." *British Journal Psychological Monograph*, No. 8 (1925).
55. MACLAURIN, DOROTHEA D. *An Experimental Investigation of Certain Alleged Relations between Physical Characteristics of the Hand and Mental Traits*. A.B. Thesis; 1921. (Filed in University of Wisconsin Library.)
56. MARTIN, E. M. "An Aptitude Test for Policemen." *Journal of Criminal Law and Criminology*, Vol. 14 (1923), pages 376-404.
57. MAY, MARK A. "Predicting Academic Success." *Journal of Educational Psychology*, Vol. 14 (1923), pages 429-440.
58. MOHR, GEORGE J., and GUNDLACH, R. H. "The Relation between Physique and Performance." *Journal of Experimental Psychology*, Vol. 10, No. 2 (1927), pages 117-157.
59. MUSCIO, B. "Motor Capacity with Special Reference to Vocational Guidance." *British Journal of Psychology*, Vol. 13 (1922), pages 157-184.
60. NACCARATI, S. "The Morphologic Aspect of Intelligence." *Archives of Psychology*, Vol. 45 (1921).
61. — and LEWY-GUINZBERG, B. L. "Hormones and Intelligence." *Journal of Applied Psychology*, Vol. 6 (1922), pages 221-234.
62. *National Academy of Sciences, Memoirs of the*. "Psychological Examining in the United States Army," Vol. 15 (1921). 890 pages.
63. PATERSON, D. G., and LUDGATE, K. E. "Blonde and Brunette Traits: A Quantitative Study." *Journal of Personnel Research*, Vol. 1 (1922), pages 122-128.
64. PATTEN, E. F. "An Experiment in Testing Engine-Lathe Aptitude." *Journal of Applied Psychology*, Vol. 7 (1923), pages 16-29.
65. PERRIN, F. A. C. "An Experimental Study of Motor Ability." *Journal of Experimental Psychology*, Vol. 4 (1921), pages 24-56.
66. PINTNER, RUDOLF. *Intelligence Testing: Methods and Results*. Henry Holt & Co., New York; 1923. 406 pages.
67. PLATO. *The Republic* (translated by Davies and Vaughan). A. L. Burt Company, New York.
68. POLLOCK, HOWARD. *The Application of Vocational Ability Tests in the Hosiery Industry*. B.Ph. Thesis; 1921. (Filed in University of Wisconsin Library.)
69. PROCTOR, W. M. "Psychological Tests and Guidance of High School Pupils." *Journal of Educational Research Monograph*, No. 1.
70. RUCH, G. M., and STODDARD, G. D. "Comparative Reliabilities of Five Types of Objective Examinations." *Journal of Educational Psychology*, Vol. 16 (1925), pages 89-103.
71. — and DEGRAFF, MARK H. "Corrections for Chance and 'Guess' vs. 'Do Not Guess' Instructions in Multiple-Response Tests." *Journal of Educational Psychology*, Vol. 17 (1926), pages 368-375.
72. — and KOERTH, W. "'Power' vs. 'Speed' in Army Alpha." *Journal of Educational Psychology*, Vol. 14 (1923), pages 193-208.

73. RUGG, H. O. "Is the Rating of Character Practicable?" *Journal of Educational Psychology*, Vol. 12 (1921), pages 425-438, 485-501; Vol. 13 (1922), pages 30-42, 81-93.
74. SCOTT, W. D., and CLOTHIER, R. C. *Personnel Management*. A. W. Shaw Company, Chicago; 1923. 643 pages.
75. SHELDON, W. H. "Morphologic Types and Mental Ability." *Journal of Personnel Research*, Vol. 5, No. 11 (1927).
76. SHERMAN, ELSIE B. *An Experimental Investigation concerning Possible Correlation between Certain Head Measurements and University Grades*. A.B. Thesis; 1923. (Filed in University of Wisconsin Library.)
77. SNOW, A. J. "Tests for Chauffeurs." *Industrial Psychology*, Vol. 1 (1926), pages 30-45.
78. SOMMERVILLE, R. C. "Physical, Motor, and Sensory Traits." *Archives of Psychology*, Vol. 12 (1924), pages 1-108.
79. SPEARMAN, CHARLES. "'General Intelligence' Objectively Determined and Measured." *American Journal of Psychology*, Vol. 15 (1904), pages 201 ff.
80. ——— *The Abilities of Man*. The Macmillan Company, New York; 1927. xxxiii + 415 pages.
81. STARCH, DANIEL. "A Scale for Measuring Handwriting." *School and Society*, Vol. 9 (1919), pages 154-155, 184-188.
82. ——— *Educational Psychology*. The Macmillan Company, New York; 1919.
83. STENQUIST, JOHN L. "Measurements of Mechanical Ability." *Teachers College Contributions to Education*, No. 130. Columbia University, New York; 1923. 101 pages.
84. SYMONDS, P. M. "Variation of the Product-Moment (Pearson) Coefficient of Correlation." *Journal of Educational Psychology*, Vol. 17 (1926), pages 458-469.
85. TAYLOR, F. W. *The Principles of Scientific Management*. Harper & Brothers, New York; 1911. 144 pages.
86. Terman, LEWIS M. *The Measurement of Intelligence*. Houghton Mifflin Company, Boston; 1916. 362 pages.
87. THORNDIKE, E. L. "Early Interests: Their Permanence and Relation to Abilities." *School and Society*, Vol. 5 (1917), pages 178-179.
88. ——— "Intelligence and Its Uses." *Harper's Magazine*, Vol. 140 (1920), pages 227-235.
89. ——— "The Permanence of Interests." *Popular Science Monthly*, Vol. 81 (1912), page 449.
90. ———, Terman, LEWIS M., and Others. "Intelligence and Its Measurement: A Symposium." *Journal of Educational Psychology*, Vol. 12 (1921), pages 123-212.
91. THURSTONE, L. L. "A Scoring Method for Mental Tests." *Psychological Bulletin*, Vol. 16 (1919), pages 235-240.

92. THURSTONE, L. L. "The Mental-Age Concept." *Psychological Review*, Vol. 33, No. 4 (1926).
93. TOLLEY, H. R., and EZEKIEL, M. J. B. "A Method of Handling Multiple-Correlation Problems." *Journal of American Statistical Association*, Vol. 18 (1923), pages 995, 997.
94. TOOPS, HERBERT A. "Tests for Vocational Guidance of Children Thirteen to Sixteen." *Teachers College Contributions to Education*, No. 36. Columbia University, New York; 1923. xii + 159 pages.
95. ——— "Trade Tests in Education." *Teachers College Contributions to Education*, No. 115. Columbia University, New York; 1921. 118 pages.
96. VITELES, MORRIS S. "Research in Selection of Motormen." *Journal of Personnel Research*, Vol. 4 (1925), pages 100-115, 173-199.
97. VOELKER, PAUL F. "The Function of Ideals and Attitudes in Social Education." *Teachers College Contributions to Education*, No. 112. Columbia University, New York; 1921. 126 pages.
98. WEAVER, ANDREW T. "Experimental Studies in Vocal Expression." *Journal of Applied Psychology*, Vol. 8 (1924), pages 23-51, 159-186.
99. WEBB, E. "Character and Intelligence." *British Journal Psychological Monograph* (1912).
100. WECHSLER, DAVID. "The Measurement of Emotional Reactions: Researches on the Psycho-Galvanic Reflex." *Archives of Psychology*, Vol. 12, No. 76 (1925), pages 5-181.
101. WHIPPLE, G. M. *Manual of Mental and Physical Tests* (2 volumes). Warwick & York, Inc., Baltimore; 1915. 688 pages.
102. WHITNEY, F. L. "Predicting Teaching Success." *Journal of Educational Research Monograph*, No. 6. 85 pages.
103. WOOD, BEN D. *Measurement in Higher Education*. World Book Company, Yonkers-on-Hudson, New York; 1923. xii + 337 pages.
104. YERKES, R. M., BRIDGES, J. W., and HARDWICK, R. S. *A Point Scale for Measuring Mental Ability*. Warwick & York, Inc., Baltimore; 1915. (Revised in 1923 by R. M. Yerkes and Josephine Curtis Foster.)
105. YULE, G. UDNY. *An Introduction to the Theory of Statistics*. J. B. Lippincott Company, Philadelphia. Fourth edition, 1917. 382 pages.

INDEX

- Abilities, 215
 - levels of, 215
 - significant law of, 215
- Analysis, 285-302
 - of assembling gun parts, 288
 - of free-hand drawing, 300-302
 - of occupation, 285 ff.
- Anatomical signs, 111 ff.
 - claims of, 112-113
- Anderson, L. D., 122
- Application of mathematics to Galton's idea, 10
- Aptitude:
 - criteria, types of, 375-376
 - action, 375
 - product, 375
 - subjective impression, 376
 - derivation of, 156
 - determiners, 195
 - differences, 21-22
 - individual, 21
 - law of individual, 22
 - striking examples of, 21-22
 - trait, 21
 - measured by means of batteries, 42-43
 - specific, 19
 - prediction scale, 156-157
 - prognosis, definition of, 2
 - of salesmen, 40
 - scope of non-academic, 20
- Army tests:
 - Alpha and Beta, 17-18
 - distinction between the two, 17
 - Alpha, 78, 173, 177, 178, 217, 308, 350
 - scope of, 17-18
 - as pencil-and-paper tests, 74
 - speed-power indication, 97-98
 - value of, during war, 18
- Attenuation, 231-232
 - correction for, 242 ff.
- Barton, E., 211
- Batteries:
 - choosing tests of, 255 ff.
 - construction of, 281-285
 - correlation of, with aptitude, 13
 - efficiency of, 264-267
 - final composition of, 254 ff.
 - influence of successive tests on, 257-258
 - number of tests advisable, 261 ff.
 - possibility of, 262-263, 265
 - practical yield of, 275-276
 - reasons for low yield of, 276-278
 - single typing, 44
 - validity of, 244-245
 - yield of, 258-259
- Bergson, H., 147
- Binet, A., 13, 14, 16, 147, 163
- Binet-Simon Tests:
 - combination of, 177
 - description of, 13-14
 - development of, 14
 - ease of determining mental age by, 16
 - weakness of, 16
 - IQ explained, 16
 - mental age and IQ, 15, 163-164
 - revision of, 15
 - sample of, 15
 - significance of, 15
- Blackford, K. M. H., 124
- Book, W. F., 291-292
- Brigham, C. C., 306, 308, 332, 367
- Bridges, J. W., 190
- Brown, L. E., 149
- Burt, H. E., 307
- Cartometer, 352-353 (illus.)
- Cattell, J. McKeen, 7-8
- Chance factors, 191 ff.
- Character judging, 121

- Character judgments, 119-120
 relation of, to aptitude, 119
 correlation of, with aptitude, 120
- Character traits, 149-150
- Charts:
 talent, 174
 temperament, 175
- Checks for r 's, 434-439
- Chirognomy, 145
- Cleeton, G. U., 124
- College Entrance Examining Board,
 90, 158
- Combination of test errors, 448
- Computation, preliminary, 423, 425
- Correlation coefficient:
 among overlapping determiners,
 216
 among various complex motor
 tests, 214
 among various motor tests, 211-
 212
 among various tests in Army
 Alpha, 209
 analysis of, 10-13
 application of, to tests, 9
 author of, 9
 between character and hand-
 writing, 149-150
 between circulatory evidence and
 voluntary behavior, 155
 between circulatory phenomena
 and academic success, 153
 between head measurements and
 grades, 134
 between intellectual and motor
 activity, 210
 between physiognomic traits and
 character traits, 130
 between pooled judgments, 117,
 123
 between proportions of hand and
 character traits, 146
 between ratings, 126
 between repetitions of card sort-
 ing, 240
 between several variables of Nac-
 carati, 143
 between true measure and fallible
 measure, 239
 first used, 9-10
 for perfect prediction, 264
 formula of, 423-424
 negative, 247
 none whatever, 246
 number required, 426
 partial, 249 ff.
 relation to forecasting efficiency,
 273
 size of, 227-228
 spurious, 249
 table, 120
 true, 242-243
 unreliability of, 226-227
 with criterion, 265
- Correlation work sheet, 427 ff.
 example of, 428
- Correlation yields, table of, 451-453
- Crelle, 429
- Dalton, J., 38
- Degraff, M. H., 315
- Determiners:
 hypothetical arrangement, 222
 of efficiency, 218
 relation between criterion and
 test, 241
 typical situation, 277 ff.
 weighted, 225
 weighted in two repetitions of the
 test, 233
- Differences in means and spreads,
 395
- Dimensions of head, 131-135
- Dispersion, 470-472
 of estimates, 470
 narrowed, 470
 shrinkage of, 472
- Distribution, normal, 168 ff.
- Dollinger, M., 190
- Downey, J., 108, 308, 322

- Early experimenters, 7-9
 Ebbinghaus, 89
 Edison, T. A., 21
 Efficiency yield of tests, 272-275
 low forecast of, 275
 range of, 274
 Error of estimate, 270
 correlation of, 270
 explanation of, 270
 Evans, A. L., 127
 Experimental data, 440
 Ezekiel, M., 436, 457
- Facilitating table, 484-486
 Factor theory:
 diagram of, 196
 extreme specific, 198
 general *vs.* specific, 195-196
 group factor, 203 ff.
 group diagram, 204
 two-factor, 197
 weighted group factors, 219
 Fernald, M. R., 108
 Forecasting efficiency, 267-268
 coarseness of criterion scale, 267
 index of formula for, 268
 rank of, 267
 Forecasts, 468
 Formulas:
 for R , 467
 for prediction, 464-465, 472-473
 for prediction of school marks, 476
 Thurstone, 444-445, 446-447
 Franklin, E. E., 191
 Free-hand drawing, 453-479
 moments of final battery for, 478
 project, 453 ff.
 table of r 's for, 454
 table of weights for, 479
 Freyd, M., 206, 208
- Galton, F., 7-8
 Garnett, J. C. M., 199-200
 Gauss, 461
- General intelligence, definition of, 62
 Gilbreth, F. B., 290
 Goddard, H. H., 15
 Goldsmith, O., 38
 Grant, U. S., 37-38
 Graphic presentation:
 critical score, 158
 dice throws, 30
 distribution of stature, 24-25
 ideal distribution curve, 31
 individual differences, 23-25
 pedagogical efficiency of grade teachers, 28
 physical growth, 165
 picture-completion test, 166
 poundage output of operators, 27
 rating scales, 407-409
 relation between r and forecasting efficiency, 272-274
 salaries of office boys, 26
 self-administering tests, 172
 transmuted scores, 29
 Graphology, 14, 147
 Gross determiners of success:
 capacity *vs.* industry, 184-185
 relative importance of, 186-187
 chance factors, 191-193
 generalization *vs.* specialization, 194-195
 interests of individual, 189-191
 various factors in success, 193
 Group tests:
 best example of, 93
 involving apparatus, 93 ff.
 Otis group tests, 16-17
 similarity of trial, 342-343
 size of trial group, 341
 Gundlach, R. H., 142
- Heibredner, E., 144
 Henmon, V. A. C., 78
 Hierarchical order of Spearman, 196
 Hollingworth, H. L., 115, 307

- Hollingworth, L. S., 40
 Holmgren, 104
 Houdini, H., 39
 Hull, C. L., 42-43, 148
 Hull correlation machine, 213, 487-489
 Intelligences, definition of, 62
 IQ, 159
 Judgments, validity of, 122-125
 Kelley, T. L., 187
 Kenagy, H. G., 123
 Kingsbury, F. A., 307
 Knight, F. B., 124
 Koerth, W., 97
 Kornhauser, A. W., 307
 Kraepelin, E., 9
 Kreisler, Fritz, 157
 Kretschmer, E., 140
 Kymograph, description of, 79
 La Place, P. S. de, 37
 Larson, S. C., 390
 Leibnitz, G. W. von, 21
 Limp, C. E., 33-35, 42-44, 181
 Link, H. C., 74, 158, 288, 307, 345, 378
 Ludgate, K. E., 123
 MacLaurin, D., 145
 McCabe, F. E., 115
 McCall, W. A., 169
 McFarlane, M., 214-215
 Martin, E. M., 406
 May, M. A., 187-189
 Mental age and IQ, 163
 Miller, W. S., 90, 177
 Miniature tests, 67-69, 71-72
 ingenious example of, 69
 value of, 71-72
 Wisconsin, description of, 67-68
 Mohr, G. J., 142
 Montgomery, R. B., 148
 Motion study. *See* Gilbreth
 Mueller, F. G., 105-106
 Multiple-regression equation, 179-183, 269, 471-481
 applied to problems, 269, 480
 disadvantages of, 182
 example of, 181
 in terms of school work, 471
 range of, in aptitude forecasting, 471
 Münsterberg, H., 9, 69
 Muscio, B., 210, 214
 Myerson, 109
 Naccarati, S., 142-145
 Napoleon I, 37
 Newton, I., 21
 Normal curve of distribution:
 curve of error explained, 26
 ideal distribution curve, 31
 point of inflection, 27
 Odell, 159
 Omnibus test, 64-66
 advantages of, 66
 example of, 65
 explanation of, 64
 limitations of, 66
 Orleans, J. S., 53
 Otis, A. S., 17-18, 84, 170, 173, 177, 319
 Paterson, D. G., 123
 Patten, E. F., 53
 P. E., 227-231
 Pearson, Karl, 10, 137, 385, 422
 Percentiles, 161-163
 explanation of, 161
 table of, 162
 Perforated paper tape, 488
 Perrin, F. A. C., 209, 211, 214
 Physiognomy, 113
 Pintner, R., 109
 Planimeter, polar, 371
 Plato, 5-6
 Poisson, 230

- Pollock, H., 27
 Preparations for testing, 358-359
 Pressey, S. L., 96
 Product-moment coefficient, 422
 Profile, measuring device for, 127-129
 Psycho-galvanic reflex, 76
 Psychograph, 173-176, 293-294
 individual, 173
 interpreted, 176
 job, 293-294
 Psychology, differential, 22

 Ranking, 115, 416-417
 table of, 416-417
 Ratios, 33-37
 Reliability coefficient, 231-234
 effect of irrelevant factors on, 234
 effect of repeating test on, 234
 Renan, E., 147
 Retention of tests in batteries, 421
 Rich, G. J., 151
 Ruch, G. M., 97, 238, 314-315

 Salesmen, blond *vs.* brunette, 123 ff.
 Scales:
 Binet, 61
 criticism of, 164
 coarseness of criterion, 267
 5-point, 408
 linear, 384
 normal distribution, 168
 qualitative, 411
 range of group variability, 166 ff.
 variety of rating, 401 ff.
 vocational aptitude efficiency, 48
 Scores:
 AI, 171
 coarse descriptive, 160
 conversion of, 396-399
 criterion, 269
 criterion score, how obtained, 341
 criterion score with test score, 283 ff.
 estimated criterion, 482
 graphic method of locating, 158
 IB, 170
 IQ, 170
 T, 170
 table of, 160
 Scoring, 354-356, 376-378
 criteria, 376 ff.
 Link's method, 378
 stencils, 354-356
 Scott, W. D., 40
 Seashore, C., 60, 93, 173
 Shakespeare, W., 21
 Sheldon, W., 144
 Sherman, E., 131
 Simon, Th., 163
 Smedley dynamometer, 101
 Snow, A. J., 287-288
 Social aggressiveness, 151
 Sommerville, R. C., 137, 141, 143
 Spearman, C., 150, 195-201
 Spearman's coefficient, 150
 Spearman-Brown formula, 236-237, 313-314, 418
 Standard deviation or S. D., 28-31
 graphic presentation of, 31
 value of, 29
 Stanford Revision, 15, 304
 Starch, D., 32-33, 411
 Starr, H. E., 152
 Stenquist, J. L., 93
 Stenquist Tests, 92-96, 354
 Stimulus-response mechanism, 100 ff.
 Stoddard, G. D., 314
 Stop watch, 348
 Strader, U., 416

 Tables, Barlow, 430
 Taylor, F. W., 289
 Terman, L. M., 15, 60, 170, 173, 305
 Tests:
 A-, example of, 102-103
 abstract traits, 70
 achievement *vs.* trade, 50
 analogies, 87
 apparatus *vs.* non-apparatus, 72

Tests (*Continued*)

aptitude signs *vs.* aptitude, 111
 aptitude *vs.* proficiency, 50
 arithmetic, 85
 Army. *See* Army Alpha and Beta
 auditory memory span, 80-81
 best answer, 86
 Binet-Simon, 13-16
 Bogardus, 210
 cause and effect, 90
 character and temperament, 107
 circle completion, 333, 336-337,
 370-371
 color blindness, 101
 conversion of individual into
 group, 82-84
 cube assembly, 100
 definition of, 4
 devised, first, 6-10
 difficulties encountered, 346-347
 directions, 85
 Downey slow-movement, 329-330,
 361
 Downey will-temperament, 363
 dummy, 345
 effectiveness of three similar, 313-
 314
 examples of, 4, 15
 experimental conditions for, 346
 general information, 88-89
 general interest *vs.* particular
 aptitude, 19
 general prudence *vs.* scholastic, 60
 groups with apparatus. *See*
 Group tests
 individual *vs.* group, 77
 intellectual *vs.* motor, 208-209
 joint correlation yield of, 260
 judgment of size, 366, 368-369,
 332-336
 Kent-Rosanoff, 207
 line extension, 336-339, 372-373
 mechanical information. *See* Sten-
 quist
 memory for design, 326-329

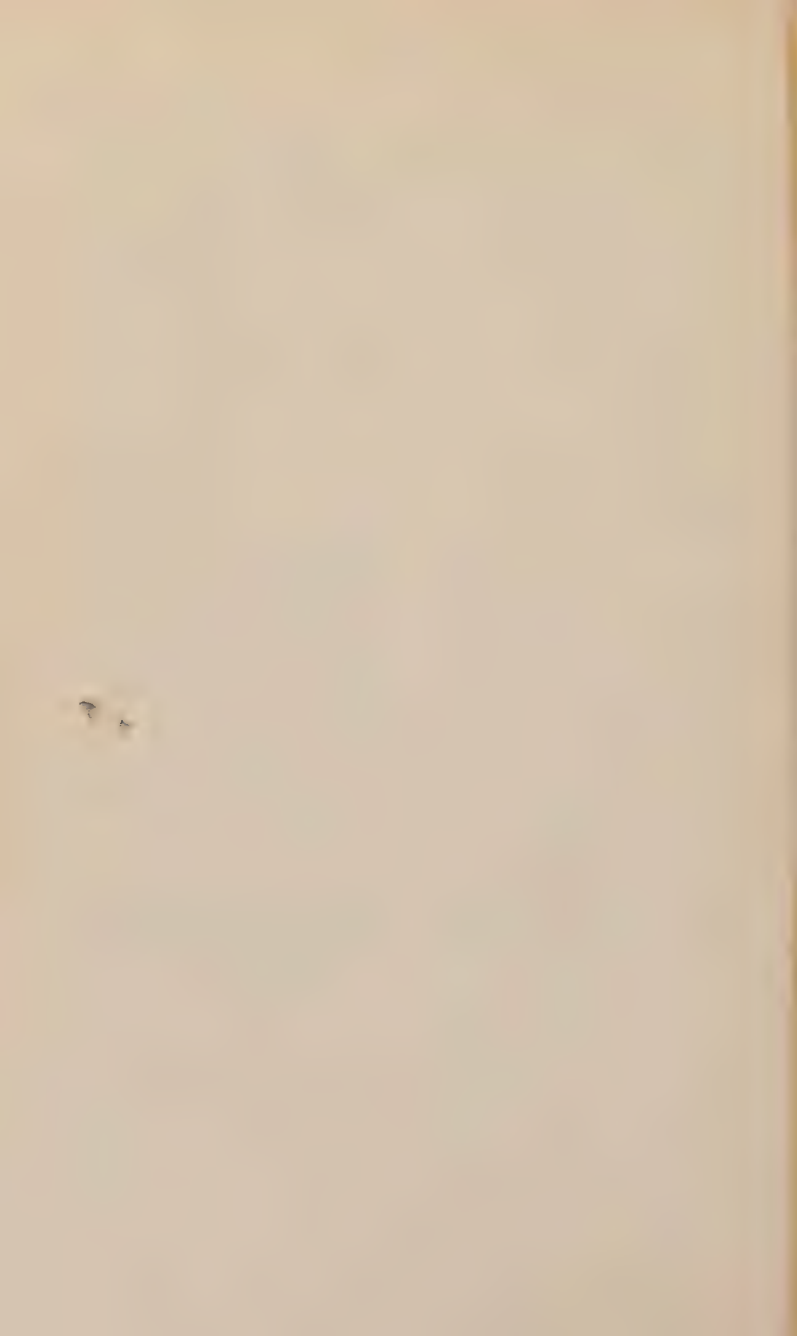
mental efficiency, 106 ff.
 Miller Mental Ability, 90
 miniature *vs.* abstract traits, 67
 motor, 211 ff.
 motor efficiency, 105 ff.
 multiple choice, 310 ff.
 National Intelligence, 60, 394
 number series extension, 87
 omnibus, 64-66
 opposites, 332, 366-367
 oral individual and trade, 54-55,
 80
 Orleans-Solomon Latin Prognosis,
 53
 peg board, 74-75
 pencil-and-paper, 74-84
 picture *vs.* performance trade,
 56-60, 92
 profile honesty, 109 ff.
 Porteus maze, 90-91
 proverbs, 89
 psychological, 7, 107
 range of, 2
 reproduction of angles, 330-331,
 364-365
 scores combined, 173-180
 scores converted into aptitude
 equivalents, 171, 173
 Seashore's musical, 60
 Seashore's pitch discrimination, 64
 self-administering, 159
 self-recording apparatus, 96
 sensory efficiency, 104
 sentence completion, 89
 shell inspection, 158
 shock absorber, 345
 simple completion, 310
 specific *vs.* general aptitude, 53
 speed *vs.* power, 96
 Stanford Achievement, 41
 Stenquist Mechanical Aptitude,
 92-94, 354
 strength of grip, 100-101
 suggestive, 109
 synonym-antonym, 86

- Terman group, 60
 testing the tests, 282, 340 ff.
 Thorndike College Entrance, 394
 three-hole, 105-106
 Thurstone technical information,
 154, 318
 time-limit *vs.* work-limit, 99 ff.
 timing of tests, 348-350
 trade, 51
 trade *vs.* aptitude, 52
 true-false, 312
 units used in batteries, 63 ff.
 weighting of, 178
 where found, 307
 will-temperament, 108, 363
 Wisconsin miniature, 67-70
 Woodworth-Wells direction, 85,
 362
 Woodworth-Wells substitution,
 107
 yields, from numbers of, 259 ff.
 Thomson, G., 198-199, 200-203
 Thorndike, E. L., 170, 189-191, 203,
 205-208, 215
 Thorndike Intelligence Examination
 scores, 142-144
 Thorndike tests for college freshmen,
 9
 Thurstone, L. L., 154, 318-319, 332
 Time study. *See* Taylor
 Tolley, H. R., 457
 Toops, H. A., 205, 313, 406
 Trait differences:
 asymmetrical development, 37
 distinction between individual and
 trait differences, 37
 investigation of blonds and brun-
 nettes, 123-126
 table of, 125
 Transmutation, 386-390, 399-400
 Types, 32, 138-141
 erroneous conception of, 32
 somatico-behavior, 138-141

 Validity of tests, 316
 Values of variables, 253
 Van Tassel, R., 153-154
 Variation, concomitant, 422
 Veeder counter, 324-325
 Viteles, M. S., 285-288
 Voelker, P. F., 109

 Weaver, A. T., 220, 394
 Wells, F. L., 73
 Weighting of tests, 457-460
 Weights, 459-463
 arrangement of, 460
 determining of, 463
 scientific method of, 457
 table of, 459
 Whipple, G. M., 104, 307, 308
 Woodworth, R. S., 73
 Work sheet, 427-443
 for r 's, 432-443
 for σ 's and square moments, 430,
 441
 Wundt, W., 6-7

 Yerkes, R. M., 17
 Yerkes-Bridges point scale, 16
 application of, to Binet scale, 16
 Yule, G. U., 23, 457





BF431 .H77

Aptitude testing



01099957 /

[illegible]

APTITUDE TESTING

